DRDC Toronto No. CR 2003-096

TRUST IN AUTOMATED SYSTEMS
LITERATURE REVIEW

by:

Barbara D. Adams and Lora E. Bruyn
Human*systems*, Incorporated
111 Farquhar St., 2nd floor
Guelph, ON N1H 3N4

Sébastien Houde
Saint Mary's University
Halifax, NS B3H 3C3

Project Manager:
Kim Iwasa-Madge
(519) 836 5911

PWGSC Contract No. W7711-017747/001/TOR
Call-Up 7747-10

On behalf of
DEPARTMENT OF NATIONAL DEFENCE
as represented by
Defence Research and Development Canada – Toronto
1133 Sheppard Avenue West
Toronto, Ontario, Canada
M3M 3B9

DRDC Scientific Authority
Carol McCann
(416) 635-2190

June 2003

Terms of release: This document contains proprietary information which is to be protected in accordance with standard business practices and is limited in distribution between participating parties. Release to third parties, without written authorization from both the originating Defence Research Establishment and the Client organization, is strictly forbidden.

© Her Majesty the Queen as represented by the Minister of National Defence, 2003

© Sa majesté la reine, représentée par le ministre de la Défense nationale, 2003

Abstract

This report reviews research literature pertaining to trust in automated systems. Based on the review, we argue that trust in automation has many similarities with trust in the interpersonal domain, but also several unique dynamics and influences. Existing research has focused primarily on trust in automation that has an executive or control function, and to a lesser extent, has considered trust in automation that is designed to present information to operators (e.g. decision aids). We maintain that although there are many similarities between trust in automation and interpersonal trust, the dynamics of trust in automation also have some distinct qualities. Several models related to trust in automation have already been developed; in this report, a comprehensive -- although still preliminary -- model of trust in military automation is proposed. Several sets of factors are likely to impact on the development of trust in automation, including properties of the automation, properties of the operator, and properties of the context in which interaction with automation occurs. The consequences of trust in automation have yet to be fully explored. Based on this review, measures and methods to study trust in automation are considered, and a program of research to study trust in automated systems is described.

Résumé

Ce rapport examine les publications dans le domaine de la recherche sur la confiance qui peut être accordée aux systèmes automatisés. Cette revue nous amène à conclure que la confiance dans l'automatisation présente de nombreuses similitudes avec la confiance dans le domaine interpersonnel, mais qu'elle présente aussi d'uniques dynamiques et subit des influences particulières. Jusqu'à présent, la recherche s'est axée principalement sur la confiance dans l'automatisation des fonctions de supervision ou de commande et, dans une moindre mesure, sur l'automatisation conçue en vue de présenter de l'information à des opérateurs (p. ex. les aides à la décision). Nous sommes d'avis que, bien qu'il y ait de nombreuses similitudes entre la confiance dans l'automatisation et la confiance dans le domaine interpersonnel, la dynamique de la première a des qualités distinctives en sus. Plusieurs modèles ont déjà été développés dans le domaine de la confiance dans l'automatisation et le présent rapport en propose un qui est complet et qui s'applique au contexte militaire. Plusieurs ensembles de facteurs sont susceptibles d'influer sur le développement de la confiance dans l'automatisation, notamment les propriétés de l'automatisation, les propriétés de l'opérateur et les propriétés du contexte dans lequel a lieu l'interaction avec l'automatisation. Les conséquences de la confiance dans l'automatisation demeurent à explorer plus avant. Cette revue des publications a servi de base à l'examen de mesures et méthodes d'étude de la confiance dans l'automatisation et un programme de recherche en vue d'étudier la confiance dans les systèmes automatisés est décrit.

Executive Summary

This report reviews the results of a keyword search of the research literature relevant to trust in automated systems. The goals of this review were to:

- Present empirical and theoretical work in the scientific and military literature relevant to trust in automation and automated systems
- Identify factors that affect the development of trust in automation
- Identify the consequences of trust in automation
- Generate a preliminary model of trust in automation
- Identify methodologies for the study of trust in automation
- Generate recommendations for a research program to explore trust in automation within the military context

The search yielded approximately 300 titles and abstracts, and resulted in the extensive review of approximately 25 articles. These articles were drawn from research in areas that included behavioural science, organizational theory, ergonomics, engineering, as well as military theory and research.

The report contains sections reviewing studies and theories relating to:

- Automation and automated systems
- Conceptualizing trust in automation
- Models related to the development of trust in automation
- Factors impacting on trust in automation
- Consequences of trust in automation
- Measures and methods of trust in automation
- Proposed research program to study trust in automation

Existing research has focused primarily on trust in automation with an executive or control function, with some consideration of automation designed to present information to operators (e.g. decision aids). Based on the review, we argue that trust in automation has many similarities with trust in the interpersonal domain, but also several unique dynamics and influences. Some models related to trust in automation have already been developed, and a model depicting a preliminary model relevant to the military context is proposed. Several sets of factors are likely to impact on the development of trust in automation, including properties of the automation, properties of the operator, and properties of the context in which interaction with automation occurs. The consequences of trust in automation beyond the simple use of automation have yet to be fully explored. Based on this review, measures and methods to study trust in automation are considered, and the features of a program of research into trust in automated systems are described.

Sommaire

Ce rapport examine les résultats d'une recherche par mots clés dans les publications de recherche portant sur la confiance dans les systèmes automatisés. Les objectifs visés par cet examen étaient les suivants :

- Présenter les travaux empiriques et théoriques dans les ouvrages scientifiques et militaires traitant de la confiance dans l'automatisation et des systèmes automatisés.
- Identifier les facteurs influant sur le développement de la confiance dans l'automatisation.
- Identifier les conséquences de la confiance dans l'automatisation.
- Générer un modèle préliminaire de confiance dans l'automatisation.
- Identifier des méthodologies d'étude de la confiance dans l'automatisation.
- Générer des recommandations portant sur un programme de recherche visant à explorer la confiance dans l'automatisation dans le contexte militaire.

La recherche a permis de relever près de 300 titres et résumés et elle a donné lieu à l'examen approfondi d'environ 25 articles. Ces articles découlent de recherches dans des domaines comprenant la science du comportement, la théorie organisationnelle, l'ergonomie, l'ingénierie ainsi que la théorie et la recherche militaires.

Le rapport contient des sections examinant les études et théories relatives aux sujets suivants:

- Automatisation et systèmes automatisés
- Conceptualisation de la confiance dans l'automatisation
- Modèles reliés au développement de la confiance dans l'automatisation
- Facteurs influant sur la confiance dans l'automatisation
- Conséquences de la confiance dans l'automatisation
- Mesures et méthodes de confiance dans l'automatisation
- Programme de recherche proposé en vue d'étudier la confiance dans l'automatisation

Jusqu'à présent, la recherche s'est axée principalement sur la confiance dans l'automatisation des fonctions de supervision ou de contrôle, avec quelque considération sur l'automatisation destinée à présenter de l'information à des opérateurs (p. ex. les aides à la décision). En nous fondant sur cet examen, nous sommes d'avis qu'il y a de nombreuses similitudes entre la confiance dans l'automatisation et la confiance dans le domaine interpersonnel, mais la première présente en sus une dynamique et des influences uniques. Plusieurs modèles ont déjà été développés dans le domaine de la confiance dans l'automatisation et nous proposons un modèle préliminaire qui s'applique au contexte militaire. Plusieurs ensembles de facteurs sont susceptibles d'influer sur le développement de la confiance dans l'automatisation, notamment les propriétés de l'automatisation, les propriétés de l'opérateur et les propriétés du contexte dans lequel a lieu l'interaction avec l'automatisation. Les conséquences de la confiance dans l'automatisation, au-delà de la simple utilisation de l'automatisation, demeurent à explorer plus avant. Cette recherche a servi de base à l'examen de mesures et de méthodes d'étude de la confiance dans l'automatisation et à la description des caractéristiques d'un programme de recherche sur la confiance dans les systèmes automatisés.

Table of Contents

ABSTRACT .... **I**

RÉSUMÉ .... **II**

EXECUTIVE SUMMARY .... **III**

SOMMAIRE .... **IV**

TABLE OF CONTENTS .... **V**

CHAPTER 1 – INTRODUCTION **1**
- 1.1 Background 1
- 1.2 Purpose 1
- 1.3 Scope 1
- 1.4 Approach to the Literature Review 2

CHAPTER 2 – METHODOLOGY **5**
- 2.1 Databases 5
- 2.2 Keywords 6
- 2.3 The Search 7
- 2.4 Selection of Articles 7
- 2.5 Review of Article 7

CHAPTER 3 – RESULTS **9**
- 3.1 Domains of Research 9
- 3.2 Structure of the Report 9

CHAPTER 4 – AUTOMATION & AUTOMATED SYSTEMS **11**
- 4.1 What is Automation? 11
- 4.2 Levels of Automation Taxonomies 14
  - 4.2.1 The Fitts' List 14
  - 4.2.2 Sheridan-Verplank Scale of Human-Machine Task Allocation 16
- 4.3 Automated Systems in the Military Context 17
- 4.4 Overview and Research Implications 18

CHAPTER 5 – TRUST IN AUTOMATION VS. INTERPERSONAL TRUST **21**
- 5.1 Defining Trust in Automation 21
- 5.2 Trust as a Psychological State 21
- 5.3 Trust as Observable Choice Behaviour 22

5.4 Need to Trust Automation 23
5.5 Referents of Trust 24
5.6 Dynamics of Trust in Automation vs. Interpersonal Trust 25
5.7 Dimensionality of Trust in Automation 28
5.8 Differentiating Trust in Automation from Other Related Concepts 28
  5.8.1 Confidence 29
  5.8.2 Use of Automation, Reliance, Cooperation 30
  5.8.3 Complacency or Overtrust 32
  5.8.4 Calibration of Trust 32
5.9 Trust in Automation in Military Contexts 33
5.10 Challenges to Trust in Automation within Military Contexts 36
5.11 Overview and Research Implications 36

CHAPTER 6 – MODELS OF TRUST IN AUTOMATED SYSTEMS 40
6.1 Muir (1994), Muir and Moray (1996) 40
6.2 Lee and Moray (1992, 1994) 44
6.3 Riley (1994) 46
6.4 Cohen, Parasuraman and Freeman (1998) 48
6.5 Seong and Bisantz (2000) 51
6.6 Madsen and Gregor (2000) 52
6.7 Kelly et al., (2001) 53
6.8 Conclusions 55

CHAPTER 7 – FACTORS AFFECTING TRUST IN AUTOMATION 58
7.1 Properties of the Automated System 58
  7.1.1 System Reliability 58
  7.1.2 System Faults 59
  7.1.3 System Components 60
  7.1.4 System Transparency 61
  7.1.5 Level of Automation 62
  7.1.6 Interactivity of the System 62
  7.1.7 Susceptibility to Tampering 63
  7.1.8 Predictability and Dependability 63
  7.1.9 Reputation of System Designer 63
  7.1.10 System Appearance 63
7.2 Properties of the Trustor 63

7.2.1 Propensity to Trust Automation 63
7.2.2 Ability to Form Mental Model of System 64
7.2.3 Trust History 64
7.2.4 Self-Confidence 64
7.2.5 Personal Work Style 65
7.2.6 Age 65
7.2.7 Cultural Influences 65
7.3 Properties of the Environment 65
7.3.1 Risk 65
7.3.2 Operational Context 66
7.3.3 Organizational Factors 66
7.3.4 Training 67
7.3.5 Task Demands 67
7.4 Research Implications 68

CHAPTER 8 – CONSEQUENCES OF TRUST IN AUTOMATION 69
8.1 Problems with Terminology 69
8.2 Trust and Use of Automation / Reliance on Automation 69
8.3 Trust and Monitoring 72
8.4 Trust and System Performance 74
8.5 Improving Trust in Automation 75
8.6 Overview and Research Implications 76

CHAPTER 9 - PRELIMINARY MODEL OF TRUST IN AUTOMATION 78

CHAPTER 10 – MEASURES OF TRUST IN AUTOMATION 84
10.1 Research Approaches 84
10.1.1 Field Study of Real or Simulated Systems 84
10.1.2 Simulators 84
10.1.3 Microworlds 84
10.1.4 Interviews 84
10.1.5 Experimental Research 84
10.1.6 Criteria, Measures and Methods 85
10.2 Trust in Automation as a Psychological State 85
10.2.1 Empirically Determined Trust in Automated Systems (Jian, Bisantz, & Drury, 2000). 86
10.2.2 Human-Computer Trust Instrument (Madsen & Gregor, 2000). 87

10.2.3 Complacency Potential Rating Scale 89
10.2.4 SHAPE Automation Trust Index (SATI) 90
10.3 Trust as Choice Behaviour 94
10.3.1 Defensive monitoring 94
10.3.2 Use of Automation 95
10.4 Overview and Considerations for Measures of Trust in Automation 95

CHAPTER 11 – PROPOSED RESEARCH PROGRAM 98
11.1 Overview and Research Considerations 98
11.2 Proposed Research Approach 99
11.3 Features of a Research Program 100
11.4 Proposed Research Approach 104
11.5 Prototypical Study 106

PRIMARY REFERENCES 108

SECONDARY REFERENCES 110

APPENDIX A – SCALE OF TRUST IN AUTOMATED SYSTEMS 114

APPENDIX B – COMPLACENCY-POTENTIAL RATING SCALE 116

APPENDIX C - SHAPE AUTOMATION TRUST INDEX (SATI V0.3) 118

CHAPTER 1 – INTRODUCTION

1.1 Background

This review stems from the work of the Command Effectiveness and Behaviour (CEB) section at Defence Research and Development Canada (DRDC) in Toronto. This section has focused previous research efforts on command and control issues, decision-making performance, stress and coping, and trust in small military teams. This review extends the scope of ongoing trust work to explore trust in automated systems. Although research and theory relevant to trust in automation is still relatively new, there is a strong foundation for beginning to understand it. This review presents a preliminary model and delineates the first stages of a future research program to study trust in automation.

1.2 Purpose

The purpose of this literature review is to develop ideas related to the empirical investigation of trust in automation. The literature review is intended to:

- Present empirical and theoretical work in the scientific and military literature relevant to trust in automation in the military context.
- Compare theories related to trust in automation with those related to interpersonal trust.
- Review existing models of trust in automation and propose a new model.
- Identify factors that affect the development of trust in automation.
- Identify the consequences of trust in automation.
- Identify measures and methodologies for the study of trust in automated systems.
- Generate recommendations for a research program to explore trust in automation in the military context.

1.3 Scope

This literature review focused on available theory and research relevant to trust in automation within the military context. Research specifically addressing trust in automation within the military context was found to be somewhat limited. Consequently, it was necessary to broaden the focus to trust in automation in a variety of domains, including engineering, trust in expert advice, and trust decisions in simulated microworlds.

Although the ultimate goal was to review trust in automation, it was deemed necessary to embrace a more global perspective in order to connect and integrate the related trust areas. As suggested by research, it has been empirically demonstrated that the human-human trust relationship is closely related to the human-machine interaction (Muir, 1994). Accordingly, this review also considered trust in automation theories in relation to theories of interpersonal trust.

This work presents existing models of trust in automation, as well as proposing a new preliminary model depicting the development of trust in automation within a military context. It should be noted that, at this stage, the model addresses only the development of trust in automation and only indirectly addresses the important consequences of trust in automation over time.

The scope of this review is limited to issues related to trust, rather than to both trust and distrust. Issues of trust and distrust, of course, are closely related, and in many ways it is impossible to address trust without at least implicitly speaking to distrust. As will be argued later, however, the dimensionality of trust (in both automation and in the interpersonal domain) continues to be problematic. Although there is no easy solution to this issue, we argue that trust and distrust are not necessarily bipolar opposites, but that they are distinct constructs in their own right, and are worthy of separate consideration. At this point, however, we have limited our discussion to trust, and violations of trust are seen as leading to lower levels of trust rather than necessarily to distrust.

1.4 Approach to the Literature Review

A two part approach to the study was employed. First, a review of the existing literature in areas close to the main topic was conducted. This overview allowed for an up-to-date analysis of existing theory and research relevant to the issue of trust in automation. In general, this review suggests that scientific understanding of the issues relevant to trust in automation within a military domain is still relatively immature and not well integrated. Nonetheless, there is a good body of research and theory that can be used to understand trust in automation in related contexts.

Recognizing the current limitations in research and theory, a rational and theoretical analysis was then employed in order to better understand issues of trust in automation likely to be relevant within the military domain. Several pertinent factors are evident from the existing literature, and have been already been explored at an empirical level. Several other factors yet to be explored, but likely pertinent, were also hypothesized to influence trust in automation.

1.5 Acronyms and Abbreviations

The following acronyms and abbreviations have been used in this paper.

Acronyms & Abbreviations	Definitions
AADM	Aided Adversarial Decision Making
ATM	Air Traffic Management
CF	Canadian Forces
COMDAT	Command Decision Aid Technology
DND	Department of National Defence
DRDC	Defence Research and Development Canada
HCT	Human Computer Trust
HF	Human Factor
NTIS	National Technical Information Service
SA	Situation Awareness
TADMUS	Tactical Decision Making Under Stress
WWW	World Wide Web

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 2 – METHODOLOGY

2.1 Databases

The following databases and sources were used to conduct the searches:

- *PsycInfo*
- *National Technical Information Service (NTIS)*
- *Social SciSearch*
- *INSPEC*
- *Ei Compendex*
- World Wide Web (WWW)

PsycInfo is a department of the American Psychological Association (APA) that offers products to aid researchers locate psychological literature. Their database is based on psychological abstracts and contains non-evaluative summaries of literature in psychology and related fields (e.g., human factors, education, business and social studies). The database contains over one million electronically stored bibliographic references with authors, titles, publication information, and abstracts or content summaries. It covers material published in over 45 countries since 1967. References include journal articles, dissertations, reports, and book chapters.

NTIS is an agency of the U.S. Department of Commerce's Technology Administration. It is the official source for U.S. government sponsored, and worldwide scientific, technical, engineering, and business related information. The database contains almost three million titles, including 370,000 technical reports from U.S. government research. The information in the database is gathered from U.S., and government agencies from countries around the world.

Social SciSearch is a database produced by the Institute for Scientific Information (ISI®). It is a multidisciplinary index for the social, behavioural, and related sciences literature, that includes all the records published in the *Social Science Citation Index*. This index provides access to all significant items (e.g. retrospective bibliographic information, authors abstract, articles, reports, etc.) from approximately 3,900 worldwide social science and technology journals. The database contains over 3,561,000 documents covering material published since 1972.

INSPEC is a bibliographic information service that provides a comprehensive index of the scientific and technical literature in physics, electrical engineering, electronics, communication, control engineering, computers and computing, and information technology. It is based on three *Science Abstract* print publications, which include *Physics Abstracts, Electrical and Electronics Abstracts, and Computer and Control Abstracts.* The database contains over 7,000,000 bibliographic records from over 3,500 journals and conference proceedings covering material published since 1969.

Ei Compendex is the computerized version of the *Engineering Index*, which provides information about the engineering and technological literature. The database covers approximately 4,500 international journals, conference papers, technical reports and books published since 1970. A total of over 4,630,000 records have been combined to create the database.

2.2 Keywords

The following set of keywords was developed for the literature search. The keywords were divided into a number of categories in order to permit an efficient assessment of the pertinent scientific literature relating trust between humans and automation. This division allowed pairing of non-overlapping keywords for the search. The "*core concept*" category was included for two reasons. First, the keywords in that category focused the search to topics directly related to trust and automation. Second, they were intended to identify any other related theoretical approaches or conceptualizations that might have been developed.

Table 1: Keywords

Category	Keyword	Related Keywords
Core Concept (1)	Trust	Adherence, Assurance, Belief, Certainty, Collaboration, Confidence, Dependence, Distrust, Doubt, Expectation, Faith, Familiarity, Fidelity, Honesty, Loyalty, Integrity, Reliance
Core Concept (2)	Automation	Adaptive Automation, Advisory System, Artificial Intelligence, Automated Control, Automated System, Automated Aid, Auxiliary Control, Decision Aid, Decision Support System, Expert System, Intelligent System
User/System Interaction	Interaction	Collaboration, Cooperation, Dependence, Expectation, Explanation, Mental Model, Ownership, Performance Attribution, Representation, Schema, Understanding, Vulnerability
Processes	Cognitive	Calibration, Cognitive Bias, Creativity, Development, Flexibility, Cognitive Strategies, Evolution, Maintenance,
	Motivational	Consequences, Cost, Expectancy, Importance, Instrumentality, Outcome Value,
	Social	Learning, Moral Obligation, Socialization, Stereotype
	Others	Abuse, Antecedents, Correlates, Disuse, Misuse/Use
Authority	Authority	Control, Decision-making, Experience, Expertise, Judgment, Monitoring, Obligation, Order, Power, Responsibility, Task Allocation,
Organisation	Organisation	Air Force, Army, Defence, Hierarchical, Industrial, Military, Navy, Virtual,
Environmental Factors	Situational Factors	Conflict, Consistency, Emergency, Error, Failure, Fault, Feedback, Predictability, Risk, Speed, Task Difficulty, Time Pressure, Workload
Measurement	Measures	Indicators, Index, Indices, Interview, Inventory, Questionnaire, Scale, Simulation, Survey, Test,
Issues Related to Trust in Automation		Attitude, Accuracy, Commitment, Culture, Goal, Involvement, Locus of Control, Multidimensionality, Values

2.3 The Search

The database searches were performed by the application of keywords from different categories into multiple combinations. First, the keywords from the Core Concept categories (e.g. Trust and Automation) were combined. The results of this pairing were examined and constituted the baseline database of "trust and automation". From this database, keywords were applied and combined from the remaining categories. The results of these second level pairings were used to determine whether the combinations needed to be redefined, in order to be more or less inclusive. When a combination generated too many references, keywords were systematically added from other categories to limit the search. Conversely, when the combination yielded too few references, keywords were dropped from the combination, or replaced with a related term. Articles obtained as the review progressed served as another source of references.

2.4 Selection of Articles

The search of the databases generated approximately 1100 titles and abstracts. A careful review and analysis was performed to eliminate non-relevant articles. The resulting 200 articles were selected and categorized by order of priority and relevance (e.g. Excellent, Good, Average, Passable, Poor). In all, 25 articles were rated as high priority and were selected as primary references. Several other articles were reviewed and used as additional resources, and are listed in the secondary reference section.

2.5 Review of Article

The 25 selected articles were analysed, and notes were taken in order to summarise the content, theories, and models applicable to trust in automation. The summarised information was combined to develop a comprehensive outline of the principal issues that would be discussed.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 3 – RESULTS

Twenty-five (25) articles were obtained for review. These articles were extracted from published journals, military and government reports, and from conference proceedings. Although these articles covered a range of research areas, the review was focused on issues relevant to trust in automation.

3.1 Domains of Research

The articles obtained for review came from the following research areas:

- Military
- Behavioural Science
- Social Sciences
- Engineering
- Information technology
- Computer Science

Unfortunately, a search for specific articles related to trust in the military context, and studies of trust in automation within this domain appeared to be somewhat limited. Most of the military literature was concentrated around the design, performance, applicability, and limitations of automated systems. However, a number of articles were selected from this domain that provided insight into the issue of trust in automation, and permitted the formulation of several conclusions.

3.2 Structure of the Report

The next chapter of this report presents an overview of the generic concept of "automation", followed by a chapter exploring the concept of trust in automation. The following section explores existing models of trust in automation evident in the literature. The factors likely to impact on trust in automation, the consequences of trust in automation, and a new preliminary model of the development of trust in automation are explored in subsequent sections. The last two sections of the report contain methods and measures that could be used to study trust in automation and a research plan for the study of trust in automated systems.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 4 – AUTOMATION & AUTOMATED SYSTEMS

In order to understand trust in automation within a military context, it is first important to explore what is meant by the term "automation". The following section discusses the definition of automation and automated systems, describes categorization frameworks for depicting varying levels of automation, and considers the types of automated systems used in military contexts.

4.1 What is Automation?

Automation has many different definitions. Although these definitions are conceptually similar, they emphasize different aspects of automation. Some definitions focus on the transfer of functions from a human operator to an automated system:

> "A device or system that accomplishes (partially or fully) a function that was previously carried out (partially or fully) by a human operator." (National Research Council, Panel on Human Factors in Air Traffic Control Automation; cited in Kelly, Boardman, Goillau, & Jeannot, 2000).

It is clear that automation involves an allocation of function from a human to an automated system. Other definitions are more specific, and specify the tasks and functions that are actually transferred from the human to the automation. Moray, Inagaki & Itoh (2000), for example, define automation as follows:

> "Automation is any sensing, detection, information-processing, decision-making, or control action that could be performed by humans but is actually performed by machine."

A very similar conceptualization of automation has been formalized in recent work. Sheridan (2002) argues that the entire scope of automation is best represented by the function that it performs. Automation refers to:

1) the mechanization and integration of the sensing of environmental variables (by artificial sensors)
2) data processing and decision making (by computers), and
3) mechanical action (by motors or devices that apply forces on the environment) or "information action" by communication of processed information to people.

The key aspects of this definition are depicted in Figure 1:

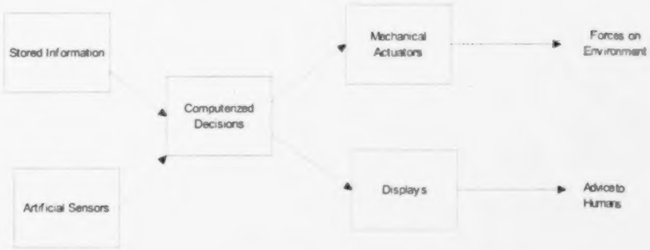


Figure 1. The scope of automation (Sheridan, 2002).

Automation can be categorized, first, as exacting forces on the environment (e.g. levers that control physical processes). This form of automation has been called executive or control automation. Examples of this type of automation include cruise control, stabilizers, thermostats and power steering (Dougherty, 2003). In terms of exerting control on the environment, simple forms of automation often include those related to process control. One prominent form of automation involves the control of physical processes. Creating paper or electricity, for example, involves product flowing though manufacturing or transportation operations. Even within this one type of automation, however, there are still varying levels of automation requiring different levels of human control. Supervisory control of such automation can range from managing manual control, to managing automatic control, to managing multiple levels of automatic control (Llinas, Bisantz et al., 1998).

Another function of automation is to present information or advice to humans. This executive or control automation aids an operator in solving problems and in making decisions. Decision aids are linked to artificial sensors in the environment, and have the role of taking information, integrating it with other forms of information, and presenting a recommendation about the best decision to be made by the operator. Fuel gauges, speedometers and clocks, and at a more complex level, expert systems and decision aids (Dougherty, 2003) are examples of this form of automation. Within each of these broad classes, automated systems perform a full range of tasks, from very simple to increasingly complex ones.

These two types of automation are most frequently found in the literature. A third type, much less frequently noted, is automation that exists exclusively for the purpose of managing other automation (Dougherty, 2003), such as flight management systems. This form of automation assists operators in managing automated systems. Automation with a management function appears to have received little attention because it has been seen as a special case of the two previous functions of automation (e.g. executive or control function vs. presenting information function).

Several distinctions can be made that further refine what automation is. First, automation is distinct from other forms of technological innovation that do not involve a change in allocation of function from humans to operators (Kelly et al., 2001). Changing a radar display with a high resolution computer display terminal is not automation, as this switch is not associated with a change in the allocation of function.

Secondly, at a societal level, it is important to note that what we consider to be automation also changes over time (Kelly et al., 2001). As Parasumaran and Riley (1997) have aptly noted:

> "When the reallocation of a function from human to machine is complete and permanent, then the function will tend to be seen simply as a machine operation, not as automation…today's automation could well be tomorrow's machine."

As a function comes to be used and accepted as a part of a system, it can become increasingly (and seamlessly) integrated into the system, and become simply a machine operation (Parasumaran & Riley, 1997). Starter motors for cars and automatic elevators, for example, perform tasks that used to require human involvement, but which are now handled independently by machines.

Thirdly, although it is convenient to speak of a system as having a given level of automation, in reality, automated systems are not necessarily unitary entities. An airplane, for example, contains many different levels of automation in its discrete subsystems. The problem with assigning a level of automation to an automated system, then, is that a given system may actually have several different subsystems, each with varying levels of automation. In this case, it is not clear how one would ascribe a single level of automation to the system.

It is also important to note that within any given form of automation, it is possible to trace the evolution of the complexity of the automation coincident with technological advances. The area of supervisory control in particular, has received a good deal of attention in this regard. Supervisory control is defined as occurring when:

> "one or more human operators are intermittently programming and receiving information from a computer that interconnects through artificial sensors and effectors to the controlled process or task environment" (Sheridan, 2002*)*

Moreover, it is also possible to depict the progression from relatively simple levels of supervisory control to much higher levels of complexity.

Figure 2 presents Sheridan's (2002) view of the progression of supervisory control.

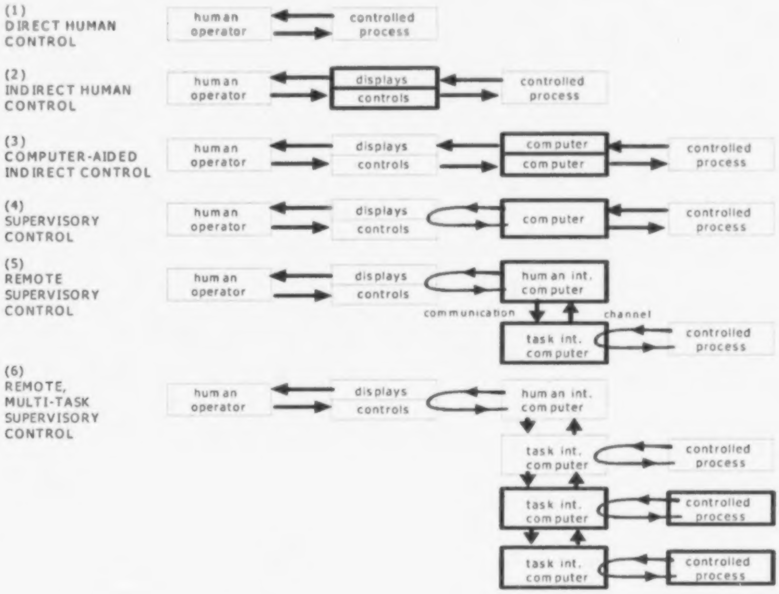

(1) DIRECT HUMAN CONTROL
human operator
controlled process
(2) INDIRECT HUMAN CONTROL
human operator
displays
controls
controlled process
(3) COMPUTER-AIDED INDIRECT CONTROL
human operator
displays
controls
computer
computer
controlled process
(4) SUPERVISORY CONTROL
human operator
displays
controls
computer
controlled process
(5) REMOTE SUPERVISORY CONTROL
human operator
displays
controls
human int. computer
communication
channel
task int. computer
controlled process
(6) REMOTE, MULTI-TASK SUPERVISORY CONTROL
human operator
displays
controls
human int. computer
task int. computer
controlled process
task int. computer
controlled process
task int. computer
controlled process

Figure 2. Levels of control from purely manual to multivariable and remote supervisory control (Sheridan, 2002).

Supervisory control, then, has progressed from a human operator exerting control directly, to the addition of displays and controls and computers, to remote supervisory control.

The relative level of automation within a given system has been formalized in approaches to categorizing levels of automation, described in the following section.

4.2 Levels of Automation Taxonomies

Automation and automated systems vary greatly in the functions that they perform. Accordingly, a number of taxonomies have been developed in order to categorize the relative levels of automation and the allocation of function to either humans or automated systems. The following section is an overview of the different taxonomies and approaches to understanding function allocation and human-machine collaboration.

4.2.1 The Fitts' List

P.M. Fitts (1951) conducted the original work on function allocation. This work was aimed at characterizing those functions that the human agent can perform better than a machine, and conversely, those functions that the machine can perform better than the human agent. The resulting classification list is summarised in Table 2:

Table 2: Fitts' List

Property	Machine	Human
Speed	- Much superior	- Lag one Second
Power	- Consistency at any level - Large constant standard forces and power available	- 2 HP for about 10 seconds - 0.5 HP for a few minutes - O.2 HP for continuous work over a day
Consistency	- Ideal for routine, repetitive or precision tasks	- Not reliable; should be monitored - Subject to learning and fatigue
Complex Activities	- Multi channel	- Single channel - Low information throughput
Memory	- Best for literal reproduction and short term storage	- Large store multiple access - Better for principles and strategies
Reasoning	- Good deductive power - Tedious to re-program	- Good indicative power - Easy to re-program
Computation	- Fast accurate - Poor error correction	- Slow, subject to error - Good error correction
Input (sensing)	- Some outside human sense range; i.e., Radioactivity - Insensitive to extraneous variables - Poor pattern recognition	- Wide range (10^{12}) and variety of stimuli dealt with by one unit - Affected by heat, cold, noise and vibration - Good pattern detection - Low signal detection - Good signal discrimination in high noise levels
Overload reliability	- Sudden breakdown	- Graceful degradation
Intelligence	- None - Incapable of goal switching or strategy switching without specific directions	- Can deal with the unpredicted - Can anticipate - Can adapt
Manipulative abilities	Task specific	- Great versatility and mobility

(Source: Backman & Digby, 1998)

Despite its staying power, the Fitt's list has been roundly criticized. One of the main criticisms of this list is that function allocation is not and should not be seen as one-time activity, which is complete once a system is designed and implemented (Moray, Hiskes, Lee, & Muir, 1995). In addition, Fitt's list does not consider the integration of the human agent with the machine. Moreover, this list of attributes has had limited impact on engineering design practice because of its level of generality, its qualitative nature, and lack of fit with engineering concepts. At best, this approach has been considered most useful at the beginning of the automation design process.

4.2.2 Sheridan-Verplank Scale of Human-Machine Task Allocation

Sheridan and Verplank (1978; cited in Levi, Moray, & Hu, 1994) were the first to introduce a formal taxonomy of automation levels to describe the modes of interaction between human and machine, called the Sheridan-Verplank Scale of Human-Machine Task Allocation (SVL). This scale was designed to organize the distribution of task allocation (responsibility) between the human and the automated agent, and to help engineers determine the appropriate level of automation for a human-machine system (Moray, Inagaki, & Itoh, 2000). Moray and Inagaki's (1999) qualitative version of this scale describes ten levels of automation in which both decisions and control pass progressively from the man to the machine (see Table 3).

Table 3. Sheridan-Verplank Scale of Human-Machine Interaction

Sheridan-Verplank 10 Levels of Human-Machine Function Allocation
1. The human does all the planning, scheduling, optimizing etc. and turns task over to computer for merely deterministic execution.
2. Computer provides options, but human chooses between them, plans the operations, and then turn task over to computer for execution.
3. Computer helps to determine options, and suggests one for use, which the human may or may not accept before turning task over to computer for execution.
4. Computer elects options and plan actions, which human may or may not approve, computer can reuse options suggested by human.
5. Computer selects action and carries it out if human approves.
6. Computer select options, plans and actions and displays them in time for the human to intervene, and then carries them out in default if there is no human input.
7. Computer does entire task and informs human of what it has done.
8. Computer does entire task and informs human only if requested.
9. Computer does entire task and informs human if it believes the latter needs to know.
10. Computer performs entire task autonomously, ignoring the human supervisor who must completely trust the computer in all aspects of the decision making.

These levels can be grouped into more generic clusters. In the first five levels, humans have responsibility for decision-making and control of the automated devices (Moray et al., 2000). Levels 5 to 7 correspond to a dynamic collaboration between the human and the machine. Finally, systems within levels 7 to 10 are considered fully automated. At the highest possible level of automation, an automated system is able to act autonomously without human intervention. Although this scale is frequently referenced, little empirical work has been conducted to determine the validity of its conceptualization of interactions between man and machine.

4.3 Automated Systems in the Military Context

Within the military context, many forms of automation are likely to be relevant. The use of computers and other automation has become a necessary tool within the military to enhance the effectiveness of its operational systems (Mastaglio, 1999). Automation affects everything on the battlefield including combat vehicles, aircrafts, communication, weapon systems, intelligence gathering, and command and control (C2) (Tyler, 1999). As technology improves, there is an increased desire to incorporate these innovations into the design of new military equipment.

Massive effort has been directed toward research and technology aimed to ensure the advanced technology needed to execute Command and Control. As such, Command and Control efforts have become increasingly automated, as depicted in Figure 3:

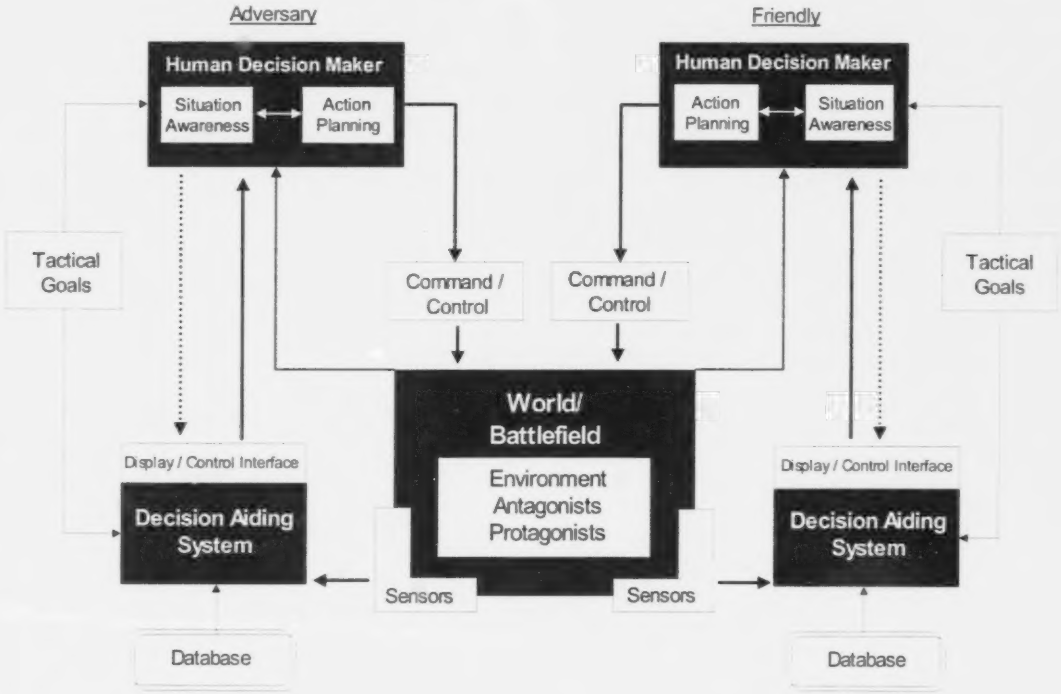

Adversary
Friendly
Human Decision Maker
Situation Awareness
Action Planning
Human Decision Maker
Action Planning
Situation Awareness
Tactical Goals
Command / Control
Command / Control
Tactical Goals
World/ Battlefield
Environment
Antagonists
Protagonists
Display / Control Interface
Decision Aiding System
Display / Control Interface
Decision Aiding System
Sensors
Sensors
Database
Database

Figure 3. Two Sided Adversarial General Model (Llinas, Bisantz et al., 1998).

This diagram shows three main nodes: the human commander (or decision maker), a data fusion system (the decision aid), and the world (as indicated in information resources). In supporting Command and Control, the sensors collect data and feed it to the information processing systems, which integrate and then display information to the human commander. Such systems support human decision making processes, which occur with many people on large-scale, hierarchical, and often geographically distributed levels. Within C2 environments, automated decision aids increasingly play a role in guiding commanders' efforts to understand the dynamic battlefield (Sheridan, 2002).

The use of automated decision aids is also prominent within other military contexts . In the Navy, for example, changes in mission objectives (e.g., peacekeeping), now require the Navy ships to patrol the littoral coastal regions of separation, in addition to conducting offensive manoeuvres. Such littoral operations represent an additional challenge for navy tactical operators (Hutchins, 1996). Navigating at close proximity to land poses the additional burden of having to accurately identify and respond to an increased number of contacts with civilian and military personnel. Operating in such a condensed and ambiguous environment requires inferences about the intent of the contacts, based on incomplete/partial information from multiple sources. Because of the limited capacity of the decision maker's cognitive resources, information gathering and processing systems have been designed to aid the decision process. The TADMUS DSS was designed by the US Navy to support decision-making aboard the AEGIS class cruiser (Hutchins, 1996). Similarly, the COMDAT, a Canadian DS and data fusion system is being designed to aid the decision processes aboard HALIFAX class frigates (Chalmers, Easter, & Potter, 2000). While a number of these automated decision support and multi-data fusion systems have been accepted, others have been criticized. These criticisms stem from the fact that automated systems can increase the decision maker's workload due to inherent system complexity, and a lack of attention to human factors during the systems design phase (Hutchins, 1996).

The air traffic control domain has made, perhaps, the most extensive use of automated systems. Within the military context, use of unmanned air vehicles (UAV's) for reconnaissance and weapon delivery is also increasingly prominent (Sheridan, 2002). Advances in electronic and computer technology have transformed the method by which pilots fly military airplanes. The cockpit control allocation has switched from manual control to a complex dynamic between the aircraft itself and the human operator. Because of the increased complexity and cognitive demand required to fly the aircraft, the pilot is now required to perform five types of tasks: (1) aviate, (2) navigate, (3) communicate, (4) manage systems, and (5) use weaponry. Such a diversified workload requires the pilot to allocate a number of these tasks to automatic devices. Automated support systems, such as Crew Assistant Military Aircraft (CAMA) (Onken, 1999) or the Cognitive Cockpit (COGPIT) (Taylor, Howell, & Watson, 2000), are theoretical examples designed to:

1) integrate multiple forms of data gathered about the surrounding environment;
2) help the pilot assess and interpret the situation;
3) aid in the planning of different courses of action;
4) facilitate decision-making;
5) help provide for a safer and more efficient flight.

This example suggests that the military use of automation runs the gamut from relatively simple levels of automation (e.g. tracking systems, data display, process control) to automated decision aids with a high level of complexity (e.g. data fusion systems and decision aids).

4.4 Overview and Research Implications

In sum, the role played by automation in the military is likely to increase in the coming years. Automation has proven useful in improving decision-making, communication, situation awareness (SA), combat efficiency, and in reducing operational uncertainty. This increasing integration between humans and automation can result in improved capabilities, but also creates a number of challenges that need to be addressed. Increasingly complex technology will enhance performance and outcomes within the military context if this technology is actually used in the way that it is intended. For this to occur, automated systems must be judged by operators to be both useful and trustworthy. If the performance of an automated system is perceived to be poor, regardless of the quality of the design, the automation may be left aside or misused. As such, understanding the role of trust in automation in military contexts will be critical. This review attempts to integrate the scientific knowledge about trust in automation, in order to assist the Canadian Forces (CF) in determining the best approach in applying and integrating technological advancements into their existing structure.

Several research implications emerge from the material presented in this chapter, as presented below:

1. Since automated systems often have multiple levels of automation, fully understanding subjective judgements about automation (e.g. trust in automation) may require the understanding of not only the global judgement, but also the judgements of the automated system's constituent parts.
2. The issues associated with trust in the automation are likely to vary by virtue of the unique properties of the function of automation (executive/ information/management). These differences will need to be considered at every stage of research.
3. As there is a great deal of variability in automation, the issues of trust associated with a very high level of automation may not necessarily be the same as those associated with simpler forms of automation.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 5 – TRUST IN AUTOMATION VS. INTERPERSONAL TRUST

How does trust in automation differ from trust in the interpersonal domain? The purpose of this section is to consider current conceptualizations of trust in automation in comparison to trust in the interpersonal domain, with respect to several key features of trust.

5.1 Defining Trust in Automation

As noted in previous work (e.g. Adams, Bryant, & Webb, 2001), a universally agreed upon and widely used definition of trust has yet to emerge. In an influential definition from the literature on close relationships in the interpersonal trust domain, Boon and Holmes (1991) define trust as:

> "a state involving confident predictions about another's motives with respect to oneself in situations entailing risk."

A definition from organizational theory defines trust as:

> "the willingness of a party to be vulnerable to the outcomes of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party." (Mayer, Davis, & Schoorman, 1995)

Within theory and research relevant to the relationship between humans and automation, several different definitions of trust have also accrued. Trust in automation has been defined as:

> "… an attitude which includes the belief that the collaborator will perform as expected, and can, within the limits of its designers' intentions, be relied on to achieve the design goals" (Moray & Inagaki, 1999).

Madsen and Gregor (2000) define trust in a decision aid as:

> "…the extent to which a user is confident in, and willing to act on the basis of the recommendations, actions, and decisions of an artificially intelligent agent"

Although these definitions stem from very diverse domains, it is also clear that there is a good deal of consistency in the key aspects of trust. These common features are explored in more detail in the sections that follow.

5.2 Trust as a Psychological State

First, there is considerable agreement that trust is best conceptualized as a multidimensional psychological state involving beliefs and expectations relevant to the trustworthiness of the target of trust (or the trustee). These beliefs and expectations are derived from experience and interaction with the target (e.g. Jones & George, 1998).

Moreover, in both the automation and the interpersonal literature, it is also clear that trust has both cognitive and affective features. From a cognitive perspective, the process of developing trust involves receiving information and knowledge about what the target of trust is likely to do in specific situations. Over time, this information becomes increasingly elaborated into views of what this agent or person is likely to do on a consistent basis (Muir, 1994; Rempel, Holmes, & Zanna, 1985). Within the trust in automation domain, there is some evidence that cognitive processes are likely to play a prominent role in determining trust expectations. Muir (1994), for example, showed that the expectation of competence about a machine seemed to best capture what people meant when they said they trusted the machine. The extent to which automation is expected to do the task that it was designed to do, then, may have a dominant influence on its perceived trustworthiness.

In the interpersonal domain, trust is also seen as involving affective processes. Developing trust requires seeing others as personally motivated by sincere care and concern to protect our interests (Lewis & Weingert, 1985; McAllister, 1995). Trust evolves as people make emotional investments in relationships, express genuine concern for the well-being of others, and come to believe that these feelings are reciprocated. This notion of investment and mutual concern is typically somewhat less prominent in a human's relationship with automation. On the other hand, it is clear that trust in automation can be argued to have an affective component. To some extent, both our attitudes and our behaviour toward automation are predicated not just on our knowledge and beliefs about it, but on how it makes us feel. The high level of frustration evidenced with a faulty automated system, for example, is likely to impact strongly on one's trust in the system independent of one's knowledge of its actual capabilities.

Studies of trust between humans and automation (e.g., Muir & Moray, 1996) also indicate that it is possible to quantify the subjective degree of trust experienced by people, and to track meaningful changes in this state over time. In short, trust in automation can be described as a psychological state with both affective and cognitive components.

5.3 Trust as Observable Choice Behaviour

In additional to being a multidimensional psychological state, trust is also expressed as willingness to engage in behaviour relevant to an agent, and as observable choice behaviour in which people are willing to make choices that put trust into action. This is equally true of trust in automation and trust in the interpersonal domain.

As a choice behaviour, trust has been conceptualized as stemming from both rational and relational perspectives. From the rationalist perspective, the decision to trust another party is motivated by a desire to make a rational, efficient choice. More precisely, the decision to trust represents a conscious computation of advantages (gains) and disadvantages (losses) that may be experienced as a result of the decision to cooperate with another agent. Trusting automation is the product of believing that the benefits of trusting will be more profitable than the costs of trusting the automation if it is not ultimately trustworthy. Rational choice behaviour is a very frequent form of trust behaviour in the automation context.

As stated by Kramer (1999), however, this depiction of trust is too narrowly cognitive. Decisions to trust most frequently occur not in rational isolation, but in the context of social systems and social experiences of the individual. As Kramer (1999) argues,

> "Trust needs to be conceptualised not only as a calculative orientation toward the task, but also a social orientation toward other people and toward society as a whole."

Considering trust in the broader context addresses the relational perspective, in which decisions to trust, and to act on that trust are more than just the calculation of losses and benefits, but a more global contextually grounded decision. Even though the context of automation is very different from the interpersonal domain, trust in automation is also relational, in the sense that decisions to trust automation involve more than the mere rational estimation of costs and benefits. One can know all the logical reasons about why automation can be trusted, and yet have an attitude so negative that trust behaviour is precluded. Similarly, one might also show trust behaviour toward automation in large part because of contextual constraints, such as the *need* to trust because of organizational mandates to do so. In considering trust in both humans and in automation, then, trust can be understood as both observable choice behaviour and a willingness to trust that have both rational and relational aspects.

As a whole, then, in both the automation and the interpersonal domain, trust is depicted as both a psychological state, involving expectations and feelings that lead to judgements about the trustworthiness of others, and as either rational or relational choice behaviour that puts these expectations and feelings into observable action. Even across widely diverse areas of study, it is important to note that there is considerable agreement on the core features of trust.

5.4 Need to Trust Automation

The need to trust automation arises from the same antecedents that influence trust in general. At a broad level, trust becomes an issue in situations that involve risk, uncertainty, vulnerability and the need for interdependence. In our interactions with automation, these antecedents are no less prominent. Automation has the potential to make complex tasks easier, and to enable us to do more in less time. At the same time, however, we are dependent on automation to function when we need it to function. When it fails, we risk not completing the tasks that automation is able to help us perform. This risk leaves us vulnerable to negative outcomes, as the costs of not performing our designated tasks can be significant. At the same time, working with automation requires at least an implicit acceptance of this vulnerability, and leaves us open to uncertain outcomes if the automation does not perform reliably.

It seems important to ask, as some trust in automation theorists have already done, why trust in automation is an issue at all. Perhaps trust in automation is a misnomer, and amounts to nothing more than the use of what works reliably, and the dismissal of that which doesn't. As Muir (1994) notes,

> "Perhaps operators simply base their allocation behaviour upon the properties of the automation: if it works, they will use it, and if it is faulty, they will not."

Clearly, basing our use of automation on the properties of the automation does occur. When automation is reliable, for example, trust is higher and the automation is more likely to be used (Muir & Moray, 1996).

Trust in automation theorists have argued, however, that trust in automation is more than just fault detection. In fact, there is both theoretical agreement as well as empirical evidence that support the view that trust and trust behaviour are based on more than just the properties of the agent. It is clear from existing research, for example, that if automation's faults are known in advance, even after these faults occur, trust is not necessarily adversely affected (Lewandowsky, Mundy & Tan, 2000). If the properties of the automation were the sole factor in trust expectations and trust-related behaviour, this would not be the case.

The existence of individual differences in the use of automation is further evidence that something other than the sheer properties of automation influence the use of automation (Muir, 1994). If the properties of automation were the only factor in the use of automation, one might expect that people would show a very similar pattern of trust toward automation and willingness to use automation. This is not the case, and is evidence that something other than rational analysis drives trust related attitudes and behaviour.

As such, trust in automation does appear to be a valid construct, and people who trust do seem to be doing more than "moment to moment fault detection" (Lewandowsky et al., 2000). Ttrust in automation, then, can be seen as an intervening variable between the operator and the operators' use of automation.

5.5 Referents of Trust

It seems obvious that trust in an inanimate object and trust in a person should also have some unique properties. Unfortunately, most of the available trust in automation literature fails to adequately articulate how trust in automation is likely to be different from trust in humans. We would argue that as the very referent of trust (i.e. either automation or a human) varies, it is important to consider how trust as a construct is likely to differ in these two domains.

There is some evidence that the trust in the automation context is similar to trust in the interpersonal context. Work by Jian, Bisantz and Drury (2000), for example, asked people to rate several attributes in terms of their applicability to understanding the trust in general, humans, or automation. Results showed some convergence in the attributes used to understand trust generally, trust in humans and trust in automation.[1]

But, there are also a number of differences between trust in an automated system and trust in a person. As Lewandowsky et al. (2000) note:

> "…at a social level, it is unlikely that the relationship between automation and humans will be identical to that between people."

This is an understatement. The very relationship between a person and an automated system does not conform to the conventional definition of a relationship. Relationships are ideally seen as being reciprocal, with both parties making contributions to the welfare of the relationship. Both parties are typically seen as accruing benefits from being in the relationship. In the case of our relationship with automation, however, this description of the relationship often does not apply. A person's trust in automation, obviously, is unidirectional – the automation does not return the trust. Similarly, the notion of two people working together to build trust (e.g. earning each other's trust over time) is also not applicable to a relationship with automation. Automation that performs consistently well over time would not typically be seen as intentionally working to earn an individual's trust.

In the case of more simple automation, the relationship between human and automation can be very limited. If automation is needed, it has a specific role to perform, and the human is entirely in control of this automation. The term "relationship" in fact, has limited power to describe this very instrumental interaction.

With the evolution of more complex technologies, however, automated systems increasingly challenge the conventional notion of the human/automation relationship as being a unidirectional one. With more complex automated systems, humans and automation can be linked in very interdependent and interactive relationships. The concept of the "human/electronic crew", for example, is representative of the paradigm shift evident in the literature between merely using automation and working with it to accomplish one's goals. In work describing an automated pilot aiding system, for example, Reising (1985; cited in Aldern, 1995) described the ideal human and automated pilot aiding system as having:

---

[1] Although, as we note later, this conclusion is somewhat problematic, as the results are somewhat inconsistent.

> "Such intimate knowledge of how to work with each other that they function as smoothly as an Olympic figure skating pair, each anticipating the moves of the other while striving for the same goal."

This relational description is equally applicable to a human team and to the human/automation "team". As this quote attests, viewing a human and an automated system as two intelligent agents working together toward a common goal has been increasingly argued to be the reality of our interactions with automation. This analogy, of course, makes little sense in relation to the more simplistic forms of automation. As we explore in the next section, differences in the referent of trust (e.g. trust in automation vs. trust in other people) are likely to impact on the dynamics of trust in relationships with these agents. This issue is discussed in more detail in the section that follows.

5.6 Dynamics of Trust in Automation vs. Interpersonal Trust

The very process by which trust in automation develops over time is also similar to the way that trust develops in the interpersonal domain. In the interpersonal domain, trust is seen to develop over time, and typically as the product of prolonged interaction with other people (e.g., Rempel, Holmes, & Zanna, 1985, Lewicki & Bunker, 1996). The process of trust development as described in the social science literature is a complex process involving the iterative development of a complex attributional profile, in which the behaviours and motivation of the target become increasingly organized into a summary judgement of the trustworthiness of the target.

In models of trust in automation (Muir, 1994), trust is also seen to develop over time and by way of many of the same processes inherent in interpersonal trust (predictability, dependability and faith). In fact, work by Muir (1989) has clearly shown subtle shifts in trust, as a response to the properties and the performance of the automation. When the automation performed reliably, operator trust increased over time. When the automation performed unreliably, however, trust quickly declined. Importantly, varying levels of trust were also strongly related to varying levels of automation use or reliance on the automation. As trust decreased, manual control became more frequent and vice versa. Within the trust in automation literature, there is also a sense in which the hypothesis testing processes used to gauge trust in the interpersonal domain are also active. Patterns of trust and trust-related behaviour, for example, seem to parallel iterative hypothesis testing, as automation operators seemed to attempt to explore the capability of the system in order to test the ability of automation to perform properly (Lee & Moray, 1994). In this sense, trust in automation appears to be dynamic in the same way as trust in the interpersonal world.

It is important to note, however, that the very goal of the relationship with automation is often different. We do not interact with automation because of the need for personal relationships, as is the case in some interpersonal trust relationships. We interact because we have a task to perform, and we begin our "relationship" with the automation purely because it is capable of helping us perform a task. The scope of the "relationship" is often constrained to whether the automated system functions properly or not, and there sometimes no need for expectations outside of this narrow scope to be even considered. Within this context, trust can be based purely on a judgement that the automation functions predictably, and there is little need for interpretation or attribution processes.

In comparison to the interpersonal models of trust development, then, many of our relationships with automation are based solely on predictability. These expectations do not require a high level of attraction or interpretation, but can be based purely on predictable behaviour (Rempel, Holmes & Zanna, 1985). Within the trust in automation context, this kind of relationship is perhaps most likely to be the case when the operator and the system work together on a more limited basis, or when there is no need for true interactivity between the system and operator.

Other relationships with automation, however, may necessitate a more complex level of trust. In cases where the operator needs to use automation frequently or when dependence on the automation is high, the trust relationship may progress to a higher level. The behaviour of the automation comes to be interpreted at a higher level of abstraction, and more as a product of the automation's underlying qualities, rather than being isolated segments of behaviour. At this point, the automation may begin to be seen as dependable at a broader level. An operator may begin to make attributions about the probable behaviour of the automated system on a given day or at a broader level.

At the highest level of trust, the operator is willing to go beyond the available information, and makes a decision to trust automation, even in the absence of information that speaks to its capability in performing a given task. At this point, faith in the system has developed. It is also important to note here that there is some evidence that trust in automation does not necessarily develop in exactly the same way as trust in people does. More specifically, the Rempel, Holmes and Zanna (1985) model of interpersonal trust argues that trust progresses from predictability to dependability to faith. Work by Muir and Moray (1996), on the other hand, suggests that for trust in automation, faith is a better predictor of trust early in the relationship, but not late in the relationship. Nonetheless, the key factors seen to influence the development of trust are similar across these varying domains.

The extent to which the full scope of the trust development process in automation can be compared to that in the interpersonal domain depends in large part on the complexity of the automation. Few people are likely to make complex attributions about the behaviour of an automated lever or a thermostat. On the other hand, one is much more likely to make attributions about an intelligent expert system used in aiding decision making.

The fact that the complexity of automation influences the potential for trust to develop to the fullest possible extent, of course, is not a limitation in the interpersonal domain. This difference arises because there must be room in the system for attributional abstraction, for people to make attributions about the system. For this to occur, there must be what can be called "poetic licence" within the system. Less complex automated systems simply offer little room for attribution.

Just as in the interpersonal domain, the development of trust in automation will also be determined not only by the complexity of the automation, but by the need for interdependence. An operator who must use a system day after day in order to perform his job has the highest potential to develop trust in the automation, as there is more need to make predictions and attributions about the automation. If the need is more constrained, trust development may be limited to simple judgements related to predictability. Perhaps for this reason, we would argue that most relationships with automation do not progress past the stage of simple predictability, very few to more complex attributions of dependability, and very few to faith. In this sense, the kind of trust relationship that is possible is likely to be influenced by the complexity of the automated automation and by the need for interdependence. More intricate systems seem to offer more potential for a complex trust relationship.

To a limited degree, the available literature addresses some aspects of the relationship between humans and automation relationship. We argue that descriptions of the trust in automation process all too often address trust in automation in isolation of trust in the human operator of the automation. In our view, it seems wrong to think about the value of what automation produces in isolation of the human element. When an automated system gives information to human operators, this does not necessarily mean that this information will be used properly. In many cases, this information will only be useful if the operator is able to understand the information, and to bring their own knowledge to bear on the situation. In this sense, then, one's trust in the outputs provided by automation are predicated to some degree on one's trust in the ability of a human operator to understand, interpret, and integrate the information that a decision aid provides. If trust in this operator is not in place, even the most reliable and intelligent automation will not ultimately be useful.

The majority of both the empirical and theoretical literature available related to trust in automation speaks to the relatively more simple levels of automation (e.g. process control), as opposed to more complex forms of automation (e.g. automated decision aids). Nonetheless, some empirical work using relatively simple automation has explored the differences in trust dynamics when the referent is an automated system vs. a human partner (Lewandowsky et al., 2000). Participants were asked to interact with what they believed was either an automated system or with a human partner. In fact, they were actually interacting with automation at all times, in performing simulated process control tasks. In general, results showed that the dynamics of trust were very similar. When faults occurred, trust in both automation and in human partners decreased quickly.

Moreover, in two experiments, the difference between trust and self-confidence was a less powerful predictor of automation use when people believed they were interacting with humans than with automation. This difference was explained in terms of what operators believed the other human thought of their trustworthiness. After operators made errors (and hence believed that they were seen as less trustworthy by the other person), they were more likely to allocate control to the person than to take it on themselves. On the other hand, the difference between operators' trust and self-confidence did not predict the use of automation when they made mistakes while interacting with automation, presumably because their perceived trustworthiness was not relevant in their interactions with automation. This suggests that the dynamics of trust are similar in some ways, in that faults diminish trust, whether caused by automation or by another person. In another way, however, self-presentation concerns related to how trustworthy we think other people see us impacts on our behaviour only in relation to humans, and not automation.

Other work addresses trust in automation vs. trust in humans with more complex automation. Research by Lerch, Prietula and Kulik (1997), for example, explored the extent to which participants "trusted"[2] the advice given to them by an expert system when they believed that this advice was from a human or a computer. A series of financial problems were presented to participants, as well as a solution or advice. Identical advice was argued to be from either a computer or from a human. Ratings of confidence in the advice, source, and agreement with the advice were taken. In the first study, results showed that participants were more confident in advice provided by humans than by the computer, but there were no differences in agreement with this advice. Moreover, participants also seemed to attribute effort to the human experts, but not to expert systems. In a later study, participants agreed more with the advice of an expert system than the human expert, but also had less confidence in this system than in the human expert. The authors argue that agreement with the system advice is paralleled in predictability judgements, whereas attributions about the source itself are indicative of dependability judgements. Unfortunately, the inconsistent results within this work limit the conclusions that can be drawn from it. What is clear is that more complex and more intelligent automated systems may challenge conventional notions of what the human/automation relationship is, and may impact on the dynamics of trust within these relationships. We would argue that we may need to rethink our relationship with all forms of automation in order to fully understand the dynamics of trust that are likely to be at play.

---

[2] It is important to point out, however, that only ratings of confidence in the advice and in the source of the advice were taken.

5.7 Dimensionality of Trust in Automation

As noted in a previous review, interpersonal trust researchers and theorists have yet to agree on the structure and dimensionality of trust. There are two main conceptualizations of the dimensionality of trust.

Trust is commonly depicted as a continuum. In this view, trust and distrust are the opposite poles of a single trust construct. A person may vary in the degree to which they trust another person, but any decrease in trust, by definition, moves them closer to distrust. This depiction of trust is evident in both the person-based model of trust presented by Lewicki and Bunker (1996), and in more recent work focused on creating a measure of trust/mistrust (Omedei & McLennan, 2000).

Trust is also conceptualized as a single construct distinct from distrust, and as a continuum ranging from high trust to low trust (Lewicki, McAllister & Bies, 1998). As a separate construct, distrust also ranges from high to low distrust. Whereas trust is defined as "confident positive expectations about another's conduct", distrust is defined as "confident negative expectations". Lewicki et al. (1998) argue that the issue of the dimensionality of trust has yet to be adequately explored in existing trust research.

Within the trust in automation literature, the dimensionality of trust is also ambiguous. Empirical work by Jian et al. (2000), for example, argues that trust is a single bipolar construct with trust and distrust on the poles. As will be reviewed later, however, this work provides conflicting evidence about the dimensionality of trust in automation. In truth, there is currently no empirical evidence that speaks conclusively to the dimensionality of trust.

Despite this, it seems important for theorists to articulate their views and to make their argument about the dimensionality of trust in their work. We would argue that trust and distrust are two conceptually distinct constructs for several different reasons. First, our trust judgements about other people or about automation are rarely wholly internally consistent. As such, it is possible to trust an agent in one domain, but to distrust in another. This dimensionality is less consistent with the view of trust on a single bipolar scale, and more consistent with the view of trust and distrust as two distinct constructs. Even more importantly, we would argue that the factors that influence trust are not necessarily the same as the factors that influence distrust. Believing that automation is competent to perform its assigned task, for example, is likely to promote trust. On the other hand, believing that it is not competent will not necessarily promote distrust, but may lead to a less trust than would otherwise be the case. In order to make these fine distinctions, it will be important to empirically establish the dimensionality of trust and distrust.

In addition, it is important to note that the assumption of our work is not necessarily that trust (in automation or in general) is good, and that distrust is bad. Rather, we would argue that some measure of both represents the ideal situation. It is important to maintain some degree of vigilance, or as Lewicki, McAllister and Bies (1998) assert, it is necessary to maintain "a healthy balance of trust and distrust". Operators must both feel that they are able to predict what the automation is likely to do, and to still believe that it is important to maintain some level of suspicion about its performance.

5.8 Differentiating Trust in Automation from Other Related Concepts

As noted in a previous review, there is confusion in the literature related to the concept of trust in both the automation literature and the interpersonal trust literature. Two issues, confidence and cooperation/reliance are relevant to both trust domains. The issue of trust calibration has been most prominent in the trust in automation domain. Complacency also receives a good deal of attention in the trust in automation literature. These concepts are explored in more detail in the following section.

5.8.1 Confidence

In the interpersonal trust literature, the construct of confidence is often used interchangeably with the concept of trust (Mayer et al., 1995). Other theorists have argued against this, and have advocated that making a conceptual distinction between them is critical. Unfortunately, trust theorists have not always agreed on exactly what the distinction between trust and confidence should be.

Luhmann (1988) argues that confidence and trust are similar in that they both involve positive expectations that may or may not lead to disappointment. Trust differs, however, in that it involves a prior engagement on the part of a person to both recognize and accept that risk exists. Judgements of trust arise in situations in which people both recognize and accept that they are at risk and are vulnerable to negative outcomes. Confidence does not require this recognition of risk. Luhmann's argument is that trust requires the situational antecedent of risk, and confidence does not.

This distinction between trust and confidence, although cited frequently in the interpersonal trust literature, stands in opposition to another conceptualization evident in the trust in automation literature. In talking about the accuracy of predicted outcomes, Muir (1994) argues:

> "…a person who makes a prediction may associate a particular level of certainty, or confidence, with the prediction. Thus, confidence is a qualifier which is associated with a particular prediction; it is not synonymous with trust."

According to this conceptualization, then, confidence is the degree of certainty associated with a specific prediction. Moreover, the probability of a specific event occurring, Shaw (1997) argues, is based not on one's intuitions, but on reason and fact:

> "Confidence arises as a result of specific knowledge; it is built on reason and fact. In contrast, trust is based, in part, on faith. We sometimes give our trust in spite of evidence that might suggest we should feel some caution, if not outright suspicion, about relying on another."

The depiction of confidence as "specific" knowledge (in the sense that the referent is very finite) is a good one. Trust judgements can, in fact, involve the same level of either specific knowledge (e.g. the probability of "X" performing a particular behaviour that I need "X" to perform), or broader attributions about the referent (e.g. what kind of person "X" is). The key difference is the existence of the contextual features (e.g. risk and uncertainty) in the case of trust.

On the other hand, we disagree with the characterization of trust as based on faith and confidence as based on reason and fact. This characterization has even been extended, with trust depicted as being affective and confidence as being cognitive (Madsen & Gregor, 2000). Again, we take issue with this conceptualization. Trust is not always based in faith, but can be based on reason and fact as well. Indeed, trust development is argued to begin as people observe the behaviour of the target with little interpretation or need for attributional abstraction. They simply observe the behaviour and the "facts" presented to them. After having observed these "facts", then, when asked to make a judgement about the probability of a similar behaviour in the future, people can simply rely on the patterns of behaviour already evidenced. As trust progresses, however, a higher level of attributional abstraction, of going beyond the information given, is required when one takes a "leap of faith" (Rempel, Holmes & Zanna, 1985). As such, trust ranges from being based on "reason and fact" to being based on expectations beyond what reason and fact would dictate (e.g. on faith).

Based on this analysis, then, several distinctions between trust and confidence can be made.

1. Trust requires risk - As Luhmann (1988) has argued, even if the decision making processes are very similar, trust occurs in the presence of risk, and confidence does not. To this, we would add that as trust is frequently conceptualized as being an issue in the presence of risk, uncertainty, vulnerability and interdependency, this entire set of situational antecedents should be added to the distinction between trust and confidence.
2. Trust can have both specific and global referents - The referents of confidence and trust are different. A confidence judgement is a discrete judgement made in relation to a specific target, in the absence of risk, uncertainty, vulnerability and interdependency. In this sense, a confidence judgement is devoid of personal self-interest and has a very specific referent. Trust judgements, on the other hand, can have either a global (e.g. the person) or a specific (e.g. the person's behaviour) referent and occur only in the presence of risk, uncertainty, vulnerability and interdependency.
3. Trust can be based on reason and fact or on faith, whereas confidence judgements are based on observable reason and fact.

Perhaps one of the reasons why trust and confidence have been so conceptually entangled is that they are sometimes highly correlated. As such, asking a person how confident they are that a specific person is likely to arrive at a scheduled meeting on time is a confidence judgement. One's judgement in this case is likely to be based on observable, rational fact, and one is likely to rely on past behaviour (e.g. "how often does this person usually arrive for meetings on time?") in order to judge the likelihood of this occurring again. In this case, there is a specific referent, an absence of personal risk or uncertainty, and the decision can be based solely on the base rate frequency of this occurrence.

On the other hand, a person's confidence that their friend will arrive on time to a mutually planned meeting is a trust judgement. In this case, however, because of the relationship with the friend, one's own outcomes (e.g. having a good time together) are dependent on the friend. Moreover, one could use a very constrained set of previous behaviours in order to judge the predictability of the friend's behaviour. In this sense, the basis of a trust judgement can be very similar to the basis of a confidence judgement. But, a broader range of information could also used to judge the likelihood of this occurring. One could make attributions about whether is person is the kind of person who is perpetually late. Secondly, while waiting for a friend, one is personally invested in the outcome of this decision (e.g. there is risk) in a way that is not the case for confidence judgments. Knowing that one's friend is likely to be late for a mutually planned meeting carries personal risks (e.g. we will not get as much time together) that are distinct from simply knowing that another person is likely to be late. As such, a confidence judgment is a discrete reason-based judgement related to the probability of a specific event occurring that occurs outside the domain of risk, and is distinct from a trust judgement.

5.8.2 Use of Automation, Reliance, Cooperation

Within the trust in automation literature, the most notable and problematic form of conceptual confusion is between the concepts of trust and reliance on or use of automation. This confusion is not unique to trust research related to automation, as noted in a previous review (Adams, Bryant, & Webb, 2001). Within the trust literature, the relationship between trust as a psychological state and trust as choice behaviour is a recurring and conceptually problematic issue. During the early stages of trust research, trust was often conceptualized purely as choice behaviour, as evidenced in the Prisoner Dilemma studies in the 1950's, exploring when people chose to cooperate with others. As trust research and theory have evolved, however, there has been a recurring call for a distinction to be made between trust-related behaviour (e.g. cooperative behaviour) and trust as a psychological state (e.g., Hosmer, 1995, Mayer et al., 1995). This argument is predicated on the observation that it is possible for people to show choice behaviour (e.g. to work cooperatively) without trusting each other. People may show trust-related choice behaviour for reasons other than trust, for example, as part of a larger strategy of competition (Hosmer, 1995), or when control mechanisms such as employers or social norms dictate the need to work together cooperatively (Mayer et al., 1995).

A conceptually similar problem occurs in the literature related to trust in automation. The assumption appears to be that trust is in play any time an operator uses automation, as evidenced in the following assertion by Muir (1994):

> "When human supervisors allow automation to control a process, we may infer that they trust that automation, to some extent at least."

In short, Muir argues that an operator must trust automation to some degree if they use it.[3] At a theoretical level, this statement seems to parallel the argument that people who cooperate with each other must have some level of trust. A worker forced to use automation day after day in order to perform his job may or may not trust the automation, but this behaviour may be the product of organizational constraints and demands, workload demands etc.[4] Making an inference that a supervisor trusts automation just because they use automation seems somewhat problematic.

This conceptual problem is also evident throughout the automation literature at a more implicit level, in research which purports to address trust in automation but which does not measure trust in automation specifically (e.g. Lerch, Prietula, & Kulik, 1997). In our view, this represents a violation of the principle that trust is a psychological state, as well as being related to specific forms of choice behaviour. Muir (1994) has argued that as an intervening variable, trust cannot be directly observed, only inferred. Indeed, the only way to know for sure whether trust is at play is to rely on measures of trust specifically geared to assess transitory psychological states. As such, we would argue that it is most difficult to make the case that trust plays a role in reliance on automation unless trust as a psychological state has been demonstrated.

There is, in fact, some evidence that a conceptual shift that addresses this problem is already underway in the trust literature. As noted in a recent review (Adams & Webb, 2003), the distinction between trust as a psychological state and as choice behaviour appears to be waning somewhat in favour of trust being conceptualised, first and foremost, as a psychological state. A recent review (Dirks & Ferrin, 2002), for example, selectively excluded articles that focused on trust as only a behaviour in favour of those that involved the conceptualisation of trust as both a psychological state and a behaviour. Other work also shows trust behaviour being relegated to a secondary role. A definition presented in a recent paper by Costa, Roe and Taillieu (2001) relegates trust behaviours to the status of a manifestation of trust as a psychological state.

> "Trust is a psychological state that manifests itself in the behaviours towards others, is based on the expectations made upon behaviours of these others, and on the perceived motives and intentions in situations entailing risk for the relationship with those others."

With this definition, it is clear that trust as a psychological state is the core of the definition, and that trust behaviours can only logically exist as the manifestation of this psychological state. If trust is conceptualised only as choice behaviour, it may be most difficult to argue that trust is the sole underlying cause of this behaviour. If trust as a psychological state is not considered, it may be difficult to be certain that trust is really at play.

---

[3] This example is the most explicit statement indicating a somewhat more implicit problem within the literature reviewed. This notwithstanding, Muir's work (1994, Muir & Moray 1996) is still the earliest and most important conceptual work toward understanding trust in automation that exists.

[4] That being said, it seems unlikely that an operator would interact with automation day after day without trust becoming an issue. Our argument is simply that even a modicum level of trust should not be inferred, just because of reliance on automation.

Unfortunately, the problems inherent in self-report measures (e.g. of psychological states) will also continue to plague future trust researchers. Whatever conceptual problems may be avoided by thinking about trust as a psychological state first and as choice behaviour second are perhaps no less daunting than capturing a person's subjective experience of trust toward automation or toward another person. Nonetheless, this distinction is an important one.

5.8.3 Complacency or Overtrust

Within the available literature related to trust in automation, complacency is conceptualized interchangeably as both the overuse of automation, the failure to monitor automation, and lack of vigilance directly due to trust in automation. A definition offered by Billings, Lauber, Funkhauser, Lyman, and Huff (1976; cited in Llinas, Bisantz, et al., 1998), for example, defines complacency as

> "Self satisfaction which may result in non-vigilance behaviour, based on an unjustified assumption of satisfactory system state."

This description of complacency as being related to non-vigilance behaviour speaks to the issue of monitoring. Complacency is also frequently associated with the construct of trust, and in fact, at several points in the trust in automation literature, the terms "complacency" and "overtrust" are used interchangeably (Llinas, Bisantz, et al., 1998).

At a conceptual level, the problem with complacency being paired with trust is that it is very possible for an operator to show complacent behaviour without this behaviour being in any way related to too much trust. Complacency can occur, for example, as the simple result of attentional focus diminishing, without trust at a psychological level being diminished. This being said, we do not dispute the notion that complacency may well be one of the critical effects of too much trust, only that equating complacent behaviour with high levels of trust is not necessarily valid in all cases. Similarly, too much trust or overtrust may well exist, but it does not necessarily and inevitably lead to complacent behaviour.

As such, we would argue that complacency should be defined at a more behavioural level without the additional inferences as to the intervening variables (e.g. trust) underlying the behaviour. In this sense, even avoiding the term "overtrust" as equivalent with complacency does not go quite far enough. It would also be helpful to make a distinction between complacency as *overuse* of automation and complacency as a lack of adequate *monitoring* of automation. In short, we argue that the construct of complacency should be disentangled from the concept of trust in automation.

5.8.4 Calibration of Trust

It is clear from the trust literature that trust in automation is an active process. As the result of interaction with a trust agent, and the outcomes of trust decisions, trust expectations and judgements are adjusted upward or downward. If trust is rewarded, it may increase and vice versa. This process has been called trust calibration.

One issue that seems problematic in the current literature, however, is how calibration of trust occurs. To what do people calibrate their trust judgements? Muir (1994), for example, argues that trust is calibrated according to the true properties of the automation:

> "If indeed supervisors' use or rejection of the automation is influenced by their trust in the automation, then their use will be optimized when their trust is at a level which corresponds to the objective trustworthiness of the automation. The process of adjusting trust to correspond to an objective measure of trustworthiness is referred to here as the calibration of trust."

From our perspective, Muir's view is problematic. In our view, there is no such thing as an objective measure of trustworthiness, any more than it might be possible to say with accuracy how trustworthy a person really is at anything more than a subjective level. One might have information relevant to performance of an automated system, and to its ability to perform its function over time (e.g. reliability). But, one can never have adequate information about the reliability of automation, as complete, objective information would make the need to trust redundant. The notion of "objective trustworthiness", then, is a misnomer. All that can really be said, perhaps, is that well calibrated trust expectations are likely to be more correlated with the actual reliability of the automation than are poorly calibrated trust expectations.

Moreover, this conceptualization of trust calibration also misses another important aspect of trust. As interpersonal trust theorists have argued, making trust judgments requires people to go *beyond* the information available to them, and to project the future state or functioning of the automation in ways that are not necessarily tied to objective facts. In making judgements to trust, one's expectations (and hence, one's calibration) will be based on much than just the actual reliability of the automation, but on a broader set of expectations and knowledge about the automation.

There is some evidence that supports this broader view. Work by Riley (1996), for example, suggests that advance knowledge about the shortcomings of automation can decrease the need for downward trust calibration. In short, when people know about when and for how long the automation will fail, their trust is unimpaired and reliance on the automation continues to be high. In short, the extent to which the system *can be predicted* is as important or more important than the extent to which the system is reliable. As such, people don't really calibrate their trust solely to the actual properties of the system, but to their knowledge of the system (with faults and without). It is important to note then, that the trust is calibrated on more than just the performance of the automation, but on the broader set of knowledge about its strengths and weaknesses. This aspect of calibration seems to be missing from most accounts of trust calibration.

5.9 Trust in Automation in Military Contexts

As noted earlier, within military contexts, automated systems are playing an increasingly important role. Ultimately, however, such systems are dependent on the human who operates them. If the human is the person to enable the decisions and to carry through with the subsequent action, understanding the operator's trust in automation is likely to be very important, as that trust can affect the final decision and the ultimate action. As Sheridan (1988) and others have argued, however, surprising relatively little attention has been paid to understanding the human aspects of performance within military environments, and even less to human/automation interactions in military contexts.

Nonetheless, Sheridan suggests seven attributes of trust, or causes of trust that can be used to understand trust within command and control systems. These attributes, and the associated definitions offered by Sheridan, are in Table 4:

Table 4: Causes of trust in command and control systems (Sheridan, 1988).

Factor in Trust	Sheridan's Definition
Reliability of the system	Repeated, consistent functioning
Robustness of the system	Demonstrated or promised ability to perform under a variety of circumstances
Familiarity of the system	System employs procedures, terms, and cultural norms which are familiar, friendly and natural to the trusting person
Understandability	The human supervisor can form a mental model and predict future system behaviour
Explication of intention	The system explicitly displays or says that it will act in a particular way (as contrasted to its future actions having to be predicted from a model).
Usefulness	Utility of the system to the trusting person in the formal theoretical sense
Dependence of the trusting person on the system	No definition offered

Although intended to address command and control systems, this framework has been influential in understanding the factors that influence trust in automation generally (e.g. Madsen et al., 2000). It is important to note that although these can be applied to all levels of automation, they would appear to be best suited to more complex levels of automation (e.g. decision aids).

A multi-phase project conducted at the Center for Multisource Information Fusion at the State University of New York at Buffalo is the most ambitious foray into understanding trust within a military context (Llinas & Drury et al., 1998; Llinas & Bisantz et al., 1998; Bisantz et al., 2000). This work focuses on aided adversarial decision-making (AADM), and on understanding the informational dependencies and vulnerabilities in using automated systems in information warfare environments. In an information warfare environment, information can be degraded or tampered with by the enemy. In the first phase of the project (Llinas, Drury et al., 1998), an analysis of the factors likely to impact on the use of decision aids was conducted. Many factors were noted as likely to impact on the use of automated decision aids, including the informational value in decision-making, cultural differences, and patterns of human error. The following conclusion guided the next phase of the work:

> "…we see the trust in automation issue as research topic rich with intellectual issues and having potentially high payoff to the military. If deep understanding about the nature of trust establishment, trust loss, mistrust, distrust, etc., can be developed, improvements in both system design and in operational procedures should be feasible, leading to high payoff in the sense of effective use of decision-aided systems even when under information attack." (Llinas Drury et al., 1998).

In order to work toward this goal, the next step was a review of the literature relevant to trust in automation (Llinas, Bisantz et al., 1998) and the creation of an empirically derived scale for measuring trust in automation (Jian, Bisantz & Drury, 2000).

One of the most interesting issues that this work raises is the potential for adversaries to promulgate low trust in automation as a tactical maneuver. Llinas and Bisantz et al. (1998) describe a scenario where adversaries may deliberately target their opponent's automation so to make the automation improperly distrusted. When using decision aids, for example, trust is predicated on the ability of the decision aid to perform an analysis of the incoming information, and to present reasonable courses of action to the operator. Tampering with the output of a decision aid, then, has the potential to make it seem unreliable or incapable. This can lead to misuse of the automation needed to sustain oneself against the adversary. In an information warfare environment, then, trust in automation is also predicated on the extent to which the automation's functioning is amenable to tampering by enemies.

Decision aids vary in the extent to which the information that stems from them can be altered. Within this kind of context, the recommendations that decision aids provide can be understood as having been intentionally degraded (e.g. sabotage), intentionally with camouflage (subterfuge), or unintentionally (e.g. occur as the result of system failure). In order to explore these ideas, a pilot study was undertaken by other researchers involved in this project (Llinas, Bisantz, et al., 1998).

This simulation involved an aircraft control task, in which participants were asked to select aircraft (identified as either hostile, friendly or unknown) on a map-based radar display and issue commands to the system about them. An information window showed non-fused data about the contact, including speed, heading, range and altitude. Based on the information available about the target, a data fusion window provided an estimate of the probability that the selected contact was friendly. Participants were required to click on unknown contacts, request either the information or the data fusion window in order to make their decision, and then identify the contact as either friendly or hostile.

There were three conditions. In the first, participants were given information that a failure was due to either sabotage, was a non-intentional system failure, or received no failure-relevant information at all (control condition). Participants completed a total of six scenarios. In the first two, the decision aid functioned normally. In the third scenario, the decision aid made an error, and the participant was notified of the fault. In the last three scenarios, the decision aid functioned normally.

Measures included accuracy in the aircraft identification task, as well as the use of decision aids. Accuracy in correctly identifying hostiles vs. friendlies was only marginally different, but there were significant differences in post-hoc comparisons between the control and hardware failure and the hardware sabotage condition. In addition, a 12-item trust scale (Jian, Bisantz and Drury, 2000) was also administered during the simulation. Performance feedback was provided at the end of the simulation.

The percentage of contacts correctly identified was first examined. This showed a very similar pattern of results across sessions. There was little difference in the use of automation in varying conditions, but the use of automation was affected by faults in general. Participants in the sabotage condition tended to use the information window less and the decision aid window more. This suggests that, contrary to prediction, participants blamed the information (e.g. relevant to speed and heading) rather than decision aid (which synthesised this information) when errors occurred. If participants believed that the decision aid had been sabotaged, it would have been more logical to reduce their use of the decision aid.

Analyses of the trust scale ratings suggests that trust ratings were very similar across all three conditions, but agreement with negatively worded statements was greater than agreement with positively worded statements. Importantly, however, there was no difference between scenarios or conditions. There are two possible ways to interpret these results. First, it is possible that trust in the automation is not dependent on the cause of the fault, that trust in a decision aid is similar even if the automation is intentionally sabotaged. Alternatively, it is also possible that the trust measure is unable to distinguish fine distinctions in trust. Despite the inability of this pilot work to show strong results, it is nonetheless the best example of studying trust in the military context. This kind of research program is closest to what would likely be desired in future trust in automation work for the Canadian Forces.

5.10 Challenges to Trust in Automation within Military Contexts

In a previous chapter, we considered the range of automation used within a military context, a context that presents several unique challenges to trust. However, it is notable that, as pervasive as the use of automation is, it is unlikely that it will ever be given complete control (i.e. that is, reach the highest possible level of autonomy) because of the high need for accountability at an organizational level.

The type of environment in which automation is expected to perform consistently and reliably is one of the main challenges to the development and maintenance of trust in automation. As Tyler (1999) argues,

> "By design, the military battlefield is a hostile environment. It is dirty, noisy, hot, cold, wet, dry and uncertain. But most of all it is lethal to men and machines. Therefore, it is essential that combat automation perform as expected. It must work – every time."

In short, in order to be trusted, automation must function reliably in the harsh physical environments faced by military personnel.

The possibility of targeted attacks on automation by enemies is also a very real threat that presents a somewhat unusual requirement on automation. Automation must not only be reliable, it must be seen as invulnerable to enemy intrusion and counter measures (Tyler, 1999; Bisantz et al., 2000).

Within these contexts, critical information related to the past performance/reliability of the automation may not always be available. Although automated equipment is sometimes used by the same people for extended periods of time (e.g. tank crews who work with and even maintain the same tank for long periods of time), this is often not the case. This, coupled with the large range of automated equipment used within military environments, will challenge the development of trust in automation. In military contexts, the fact that automation can undergo relatively low levels of use during non-operational time presents an additional challenge to trust development. The problems of maintaining older and sometimes outdated equipment, combined with a military system's reputation for automation failures (due, in part, to the sheer amount of technology used) are also likely to challenge the development of trust in automation.

Several human factors will also influence trust in automation in military settings. It also seems likely, for example, that errors and biases in human processing could play a role in military contexts in which the risks of failures are very high and failures are especially salient. In this case, even a single trust violation in a very high risk situation may hinder the ability to build trust in automation in the future. Within Canadian Forces, existing attitudes toward the use of automation also have the potential to influence trust in automation. Age of operators, for example, as well as cultural diversity and the differing levels of experience with automation will also impact on trust. Clearly, there are several challenges to understanding trust in automation within the military domain.

5.11 Overview and Research Implications

As a whole, there is a good deal of convergence in how trust in automation is conceptualised and how trust in the interpersonal domain is understood. Within the trust in automation domain, there has been considerable focus on behavioural aspects, sometimes to the detriment of thinking about trust as a psychological state. This has resulted in conceptual ambiguity (e.g. the *use* of automation being seen as tantamount with trust in the automation). In general, although there are several key differences in trust between the two domains, there is also a good deal of agreement and common understanding and usage of the term "trust". At a broad level, the trust development process is also very similar in the two domains, although the range of the development of trust in automation may be less than in the interpersonal domain. Several considerations for future research should be noted.

1. There has been little theoretical or empirical attention paid to the differences between trust in human/automation and human/human relationships. We would argue that understanding these differences will make an important contribution to understanding trust in both areas of study. As such, a key focus of this research should be the extent to which the dynamics of trust in interacting with more complex automation are compared to those in interacting with more simple forms of automation.
2. It will be important to consider trust in automation as both psychological state and as choice behaviour. We would argue, however, that choice behaviour only be considered in conjunction with a measured psychological state. In terms of the "tension" between views of trust as a psychological state, and trust as choice behaviour, a more specific review of the literature related to the attitude/behaviour relation may be helpful to be able to disentangle these issues at both a conceptual and an empirical level.
3. Trust in automation is likely to have both cognitive aspects, expressed in beliefs and expectations about the automation, as well as affective and motivational aspects, expressed in feelings and intentions toward the automation. All of these components need to be considered in efforts to measure and research trust in automation.
4. It is important to make a distinction between automation that exerts forces on the environment, and automation that provides advice to humans. The available literature, although making a theoretical distinction between them, has yet to explore this distinction empirically. This is likely to have important implications for understanding trust in automation. Is trust in both these forms of automation the same?
5. A similar issue arises in attempting to understand trust in varying levels of automation. We would argue that what is likely to distinguish trust is the degree of interpretive licence or attributional abstraction that higher levels of automation allow relative to more simple levels. This suggests that understanding trust in decision aids, for example, is not likely to be the same as trust in a vehicle. This issue should be explored at an empirical level.
6. The relationship between trust and self-confidence has been shown to be an important predictor of automation use, at least in relatively simple process control simulations. This issue needs to be explored further.
7. The dimensionality of trust in automation remains problematic. An empirical resolution of this issue is made even more important by the advancement of technology into the information warfare context, where the promulgation of distrust could be a strategic advance.
8. It will be important in any program of research to keep trust in automation and the use of automation, complacency, and confidence conceptually distinct.
9. Trust judgements are piecemeal as well as global judgements. In the case of automation, it may be possible to trust the physical hardware, but have little trust in the information that it yields, depending on the context in which the information emerges. Particularly within military contexts, for example, when adversaries could sabotage the data that a decision making aid provides, it is important to make this distinction.

10. Within the military context, the need for accountability is paramount. This may have implications for the relationship with automation, as it suggests that full and absolute control may never (and perhaps should never) be given to an automated system. From a research perspective, then, it may be possible to constraint future research efforts to no more than SVL 7 (that is, the computer does entire task and informs the human of what it has done) on the Sheridan/Verplank levels of automation categorization.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 6 – MODELS OF TRUST IN AUTOMATED SYSTEMS

There are several models depicting the development of trust in automation. This chapter summarises the existing theories and models related to trust in automation and considers their applicability to understanding trust in automation within a military context.

6.1 Muir (1994), Muir and Moray (1996)

The study of trust in automation was initiated by Muir (1994) following a realization that trust in automation was under-researched in the engineering psychology literature. Muir proposed that the study of trust in automation could be understood by adapting existing theories and models of trust from the interpersonal domain. Based on the sociological definition of trust, she developed a two-dimensional framework for studying trust in human-machine relationship by merging the dimensions of trust developed by Barber (1983) and the dimension of trust dynamics developed by Rempel, Holmes, & Zanna (1985). Muir's (1994) model is presented in Table 5.

Table 5: Muir's (1994) Two Dimensional Framework for the Study of Trust in Automation

Basis of expectation at different levels of expertise			
Dimension from Barber (1983)	Dynamic Dimension from Rempel et al. (1985) →		
Expectation	Predictability (of acts)	Dependability (of dispositions)	Faith (in motives)
Persistence			
Natural physical	Events conform to natural laws	Nature is lawful	Natural laws are constant
Natural biological	Human life has survived	Human survival is lawful	Human life will survive
Moral Social	Human and computers act "decently"	Human and computers are "good" and "decent" by nature	Human and computers will continue to be "good" and "decent" in the future
Technical competence	j's behaviour is Predictable	j has a dependable nature	j will continue to be dependable in the future
Fiduciary responsibility	j's behaviour is Consistently responsible	j has a responsible nature	j will continue to be responsible in the future

The first dimension of Muir's framework (seen in the first column in Table 5) represents three human expectations that Barber (1983) proposed are the basis for the development of trust between man and automation. Specifically, these expectations are *persistence,technical competency*, and *fiduciary responsibility* (Barber, 1983). Based on Barber's (1983) definition,

> *persistence* refers to the expectation of persistence of the natural and moral social orders.
>
> *technical competence* refers to the ability of the other agent to demonstrate role performance; everyday routine performance, technical facility, and expert knowledge.
>
> *fiduciary responsibility* refers to the expectation that the other agent will have the moral obligation to prioritize the other agent's interest before their own.

The second dimension of Muir's framework of trust in automation (seen in the top row of Table 5) represents the dynamic nature of the trust relationship as described by Rempel et al. (1985). Consistent with the Rempel et al. model of interpersonal trust, Muir (1994) argues that trust changes as a result of interactions between the human operator and the automated system. Early in the relationship, trust in an automated system is based on the *predictability* or consistency of the automation's behaviours (Muir, 1994). As the operator gains experience with the automated system, the nature of the trust changes and becomes based upon the operator's attribution of *dependability*. Prolonged experience with a system (especially in events involving risk) allows the operator to make generalizations from the specific behaviours that the system performs, to a broader set of attributions about the nature of the automation. As the highest level of trust, the operator is then able to project beyond what can be observed to a broader attribution about the operator's belief in the future dependability of the system. At this point, *faith* in the system has developed.

Muir (1994) suggested that the two dimensions in her framework of trust in automation are orthogonal counterparts. That is, the dimensions independently contribute to the development of the trust between a human and an automated system. Furthermore, this additive model of trust assumes that the three expectations of persistence, technical competent performance and fiduciary responsibility, are exhaustive and represent a linear relationship. Therefore the resulting regression model of trust takes the following form:

$$Ti = E_i\,(P_j) + E_i\,(TCP) + E_i\,(FR_j)$$

$$= B_0 + B_1X_1 + B_2X_2 + B_3X_3 + B_4X_1X_2 + B_5X_1X_3 + B_6X_2X_3 + B_7X_1X_2X_3$$

where: Ti = trust in automation

i = individual holding the expectations

j = referent (automated system)

$B_{0\text{-}7}$ = Parameters

X_1 = P (persistence)

X_2 = TCP (technical competent performance)

X_3 = FR (fiduciary responsibility)

Muir and Moray (1996) conducted a series of experiments to determine whether Muir's (1994) integrated model of trust could be meaningfully applied to trust in automation. Both experiments were conducted using a computer-controlled simulation of a milk pasteurization plant where the participants were required to control the pasteurization process. In this medium-fidelity simulation, two subsystems were automated: the pump subsystem and heating system. The pump subsystem was semi-automated, with the operator having the option to switch between manual and automatic control. The heating subsystem was manually controlled in experiment 1, and fully automated in experiment 2. The participants were required to optimize milk production by varying the balance between manual and automatic control.

The first experiment was unsuccessful in supporting a relationship between trust and use of the automation, as measured by time spent in automatic mode. The reason for this was that most of the operators engaged in extensive manual control, creating a ceiling effect. This bias toward manual control presented a problem for understanding the relationship between trust and the use of automation.

In experiment 2, the heating subsystems were programmed such that they were fully automated in order to reduce the manual control bias. The results showed a strong positive relationship between the operator's trust and use of the pump in the automatic mode. Moreover, this study showed that overall trust in the pump was affected by both the control and the display of the pump. Overall, these results show strong support for the first dimension of the model of trust in automation where the expectation of competence (e.g., flow rate) accounts for the majority of the variance, followed by the expectation of responsibility (e.g., maintain system volume). These results also offer some empirical support for the second dimension of the model. That is, the development of trust over time (i.e. from predictability to dependability to faith) accounted for a high proportion of the variance in the model. However, the data show that faith may be a better predictor of automation usage at the beginning of the relationship with the machine, as opposed to late as suggested by Rempel et al., (1985). In summary, these results suggest that the same factors influence both interpersonal trust and trust in automation, the factors may come into play at different times in trust development.

Muir (1994) also developed a qualitative model of trust in automation that defines the relationship between automation, operator's trust and predictions about the automation behaviour. It represents an extended depiction of Muir's regression model of trust (shown in Figure 4) with a focus on how operators can calibrate their trust in an automated system.

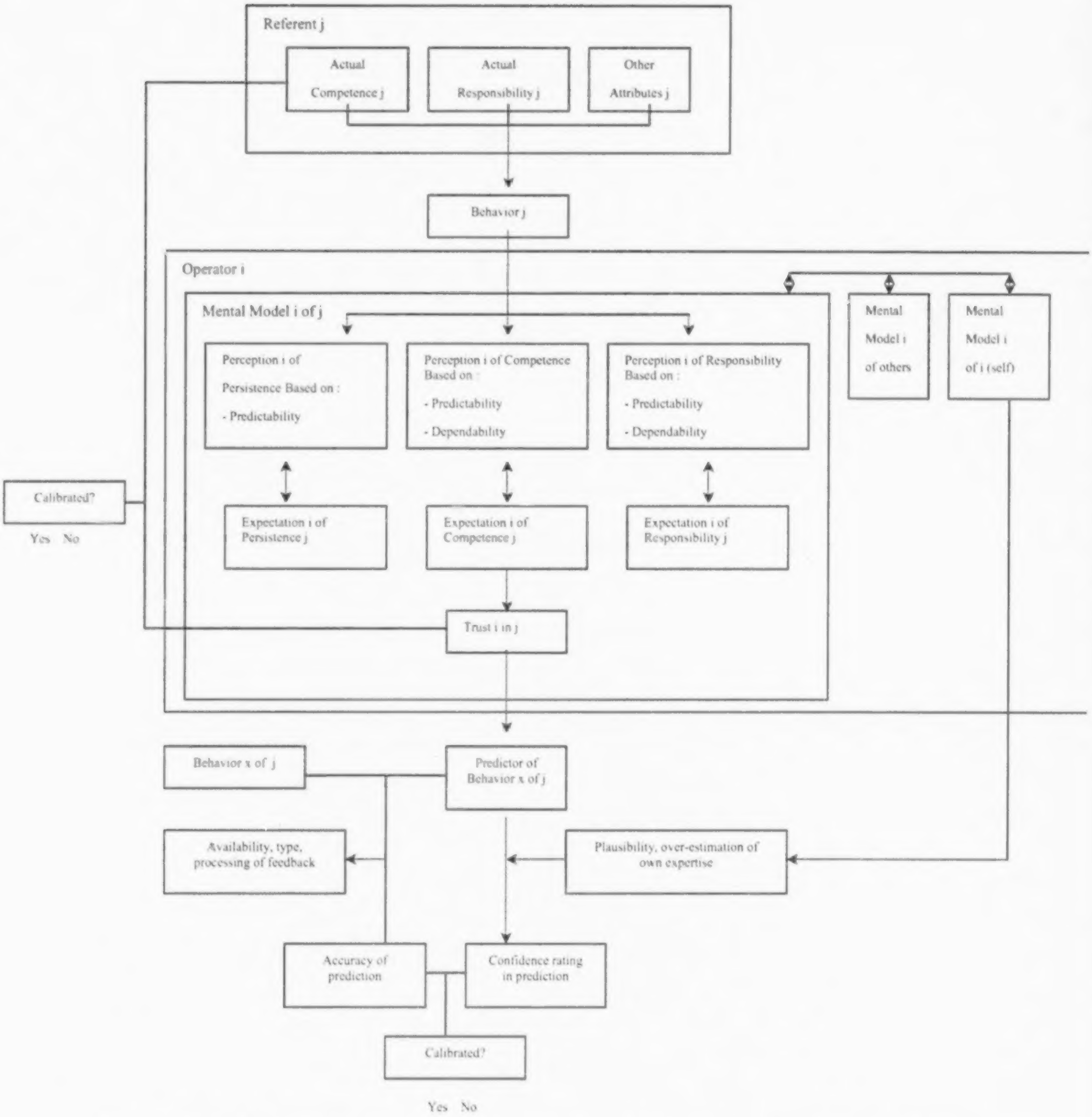

Referent j
Actual Competence j
Actual Responsibility j
Other Attributes j
Behavior j
Operator i
Mental Model i of j
Mental Model i of others
Mental Model i of i (self)
Perception i of Persistence Based on : - Predictability
Perception i of Competence Based on : - Predictability - Dependability
Perception i of Responsibility Based on : - Predictability - Dependability
Calibrated?
Yes No
Expectation i of Persistence j
Expectation i of Competence j
Expectation i of Responsibility j
Trust i in j
Behavior x of j
Predictor of Behavior x of j
Availability, type, processing of feedback
Plausibility, over-estimation of own expertise
Accuracy of prediction
Confidence rating in prediction
Calibrated?
Yes No

Figure 4: Muir's (1994) Model of the Relationship Between the Automation, the Operator's Trust, and Prediction about Automation's Behaviour.

The objective of this model was to provide a theoretical framework for interpreting and integrating future research in human trust in automation. An important component of this model is that it predicts that operators must have enough experience with the system and related problems (e.g., faults, perturbation) to develop and calibrate their trust to the specific automation.

The model developed by Muir (1994) represents the first significant contribution to the study of human trust in automation. Furthermore, this model has extensive conceptual links that have been empirically demonstrated (e.g. the link between trust and the use of automation), and is able to capture and explain a number of problems encountered with the use of automation (Moray & Inagaki, 1999). Despite its seminal contribution to understanding trust in automation, however, this model seems most applicable to the process control domain, and less applicable to understanding trust in more complex automation.

6.2 Lee and Moray (1992, 1994)

Adopting the framework developed by Muir (1989), Lee and Moray (1992) suggested a redefinition of the model of trust to include not only the dimensions described by Barber (1983) and Rempel et al., (1985), but also the dimensions of Zuboff (1988). Lee and Moray's proposed relationship of the different dimensions of trust is shown in Table 6 below.

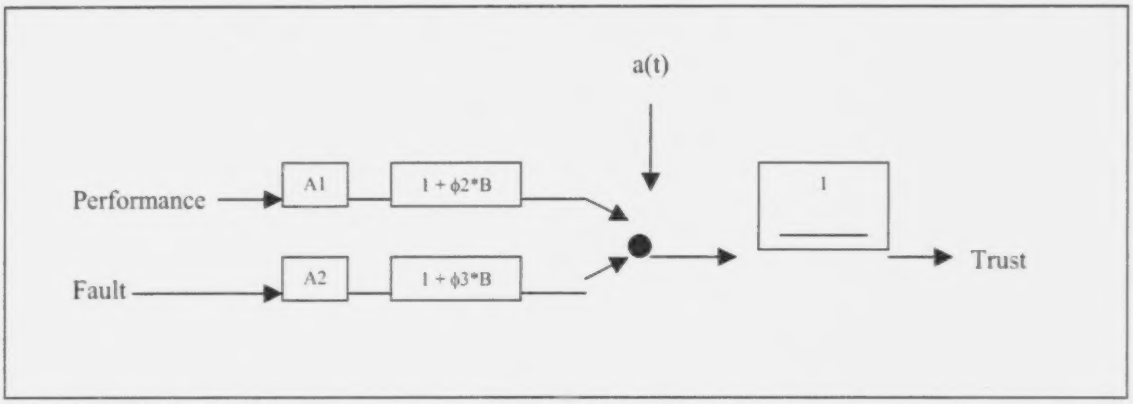
Table 6: Lee and Moray (1992) Proposed Relationship of the Different Dimensions of Trust

	Barber (1983)	Rempel, Holmes and Zanna (1985)	Zuboff (1988)
Purpose	Fiduciary responsibility	Faith	Leap of faith
Process		Dependability	Understanding
Performance	Technically competent performance	Predictability	Trial and error experience
Foundation	Persistence of natural laws		

The first additional dimension, *leap of faith*, refers to the adoption of new technology. The second, *understanding*, refers to the development of expertise or intellectual skills with the machine. Finally, the third, *trial-and-error experience*, refers to the experience with the technology over time. It is argued that these additional dimensions define the development and change in the amount of trust an operator has in an automated system. Contrary to Muir's early theoretical work (1989), Lee and Moray (1992) suggest that the dimensions of trust proposed by the three sociological theories are more complementary than orthogonal. Accordingly, the new trust model comprises four dimensions that match these three theories of trust. The first dimension represents the foundation of trust that corresponds to the persistence of natural laws defined by Barber (1983). The second dimension, performance, is characterized by one's expectations of consistency, stability, and desirability of behaviour. The third dimension, process, is based on the operator's understanding of the qualities or characteristics of the automation (e.g., algorithms, data reduction method). The final dimension, purpose, corresponds to the notion of system intent, motivation and responsibility. Specifically, it represents the designer's intention in creating the system. In sum, this theoretical framework reflects the same increasing level of attributional abstraction found in Rempel et al. (1985), as trust is argued to change from judgements related to the performance of automation to increasingly elaborated views and interpretations about it (i.e. dependability), to a more complex decision to trust automation (e.g. faith).

Lee & Moray (1992, 1994) used a simulated orange juice pasteurization plant in which participants were responsible for allocating control to either a manual or automatic process for the feedstock pump, the steam pump, and the steam heater. In order to promote automatic control (to avoid a ceiling effect), participants were also required to manually log data about the process, thereby increasing their workload. In the experiment, the feedstock pump, steam pump and steam heater experienced transient and constant faults of varying magnitude, after which operators were asked to evaluate the predictability, dependability, faith and trust in the system. The performance measure was the amount of input flow of pasteurized juice. In general, the results of the experiment suggested that when faults occurred, operator trust decreased immediately but eventually recovered. It was also determined that trust does not recover as quickly as performance, suggesting inertia in the operator's levels of trust. It was also shown that loss of trust caused by transient faults is proportional to the magnitude of the faults. That is, the larger the fault, the longer the trust recovery process. Interestingly, although fault magnitude affected trust, it was not found to affect performance. Finally, the study showed that reliance on automatic control seems to be a function of operator's self-confidence in his or her own capabilities and of the perceived system performance. Therefore, it was suggested that a pairing of trust and self-confidence might provide a better explanation for function allocation.

Lee & Moray (1992) used this data to develop causal (regression) and dynamic (time series) models of trust. The time series model, shown in Figure 5, used an Autogressive Moving Average Vector (ARMAV) to represent the dynamics of trust during fault management.



Trust (t) = ϕ_1 Trust (t-1) + A_1 Performance (t) + $A_1\phi_2$ Performance (t-1) + A_2 Fault (t) + $A_2\phi_3$ Fault (t-1) + a(t)
Where : t = time subscript, A_1 = the weighting of system performance, A_2 = the weighting of the occurrence of the fault,
ϕ_1 = Autoregressive moving average vector form time constraints, a = random noise perturbation

Figure 5: The Transfer Function of Trust Using an Autoregressive Moving Average Vector

As hypothesized by the researchers, the time series model was more proficient at uncovering the dynamics of an operator's subjective rating of trust. The regression model accounted for 53% of the variance in the level of trust, while the time series model accounted for 79% of variance. The time series model better fit the data by accounting for about a quarter more variance than the regression model. The times series model revealed that only recent events (i.e. faults) affect the operators' trust in the system and their allocation of function. This means that no more than one or two "steps backs" need to be included in the time series analysis.

Lee and Moray's time-series model represents the first attempt to model the dynamics of an operator's trust in automation and how trust is affected by properties of system faults as well as previous performance of the system. The inclusion of self-confidence as an important consideration in the use of automated systems and trust in automated systems is also important. The ability of this model to explain large amounts of the variance in trust from knowledge of the physical properties of the system is impressive. However, it should also be noted that the limited scope of the model limits the conclusions that can be drawn from it. Although this kind of model seems to predict trust in automation relatively well, it is also clear that trust in more complex forms of automation may not necessarily be as easy to predict. This work does suggest, however, that time series designs may be able to help understand trust in automation while considering a broader range of factors (e.g. levels of automation and situational factors).

6.3 Riley (1994)

Riley's (1994) model of automation shows the numerous factors influencing reliance on automation, including trust in automation. As shown in Figure 6, this model uses solid lines to indicate influences supported by empirical evidence and dashed lines to indicate hypothetical links.

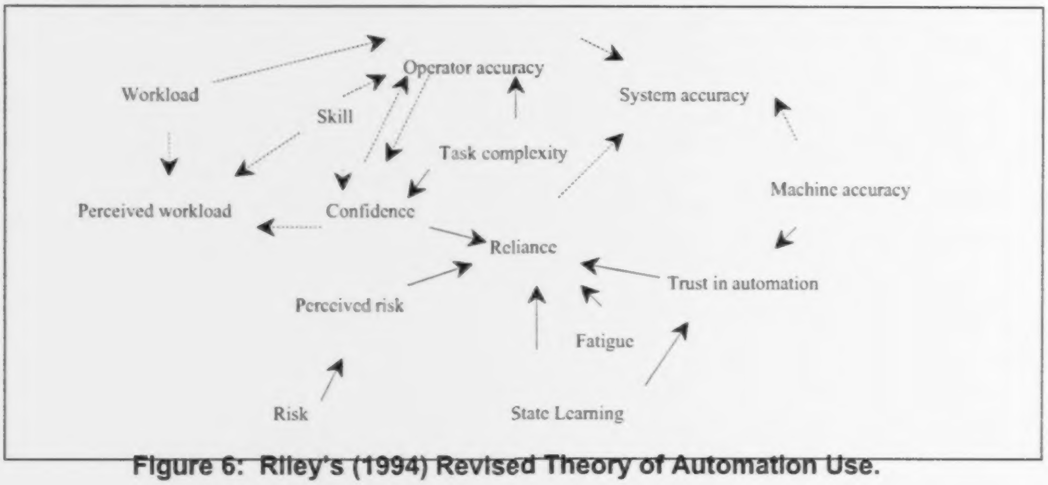


Figure 6: Riley's (1994) Revised Theory of Automation Use.

The factors influencing reliance on automation in this model were derived from a review of human psychology literature. Knowledge of the literature was used to infer and suggest specific relationships between the factors that could potentially relate to human-machine interaction.

Riley (1994) used a total of four experiments to investigate the validity of the relationships in his proposed theory of automation use. Each experiment included a computer-based task, in which participants had to decide whether or not to rely on the automation to perform the task, and a series of questionnaires. The computer-based task consisted of a simple computer game and a gambling task. The computer game was used to determine how manipulations of workload, task uncertainty and automation reliability affected automation use, whereas the gambling task provided an objective measure of risk taking, to be compared with risk questionnaire results. The objectives of the questionnaires were to determine how individual attitudes toward risk and automation affected automation use decisions and to better understand the factors that influenced one's decision to use automation.

Riley's computer-based task included both a primary and secondary task. In the primary task, subjects were required to classify randomly presented characters as a letter or a number. Subjects could assign a task to automation or retake manual control at any time by pressing the space bar. In the secondary task, subjects were required to monitor the position of a vertical line directly above a carat symbol (^) and move the line appropriately when it moved off target. This task could not be assigned to automation.

Results of Riley's (1994) four experiments demonstrated that automation reliability, task uncertainty and risk influence automation use decisions. Experiment two allowed Riley to separate out the contribution of trust in automation from uncertainty about automation states and further develop understanding of the dynamic behaviours of automation use. Conversely, results of the four experiments failed to provide a link between workload and automation use and also suggested that people are generally not good at assessing their own expertise.

Although the experiments provided some empirical support for the conceptual links, Riley's conclusions are limited by the methodology employed. They are based on laboratory experiments that did not simulate real-life mechanical systems, such as those used by Muir and Moray (1996) or Lee and Moray (1992; 1994). This may limit the generalizability of the results to real-world automation. In addition, Riley's model incorporates a tremendous number of factors that may influence reliance on automation, one of which is trust in automation. Validation of a model of this complexity and with this number of factors may prove to be very difficult, if not impossible.

6.4 Cohen, Parasuraman and Freeman (1998)

Recent research by Cohen, Parasuraman, & Freeman (1998) has challenged the work done by Muir (1989, 1994) and Lee and Moray (1994). They argue that a model of trust in automation required more clarity and theory that is applicable to training design. As a result, a computational model of trust in decision aids was developed. This model considers trust in automation in relation to arguments and uncertainty regarding system performance. The creators of this model wanted develop a probabilistic theory that 1) could capture a more differentiated concept of trust that takes into consideration the context and temporal scope of trust, and that 2) could be used to generate training scenarios based on event trees. This framework (seen in Figure 7 below) has been referred to as the Argument-based Probabilistic Trust (APT) model.

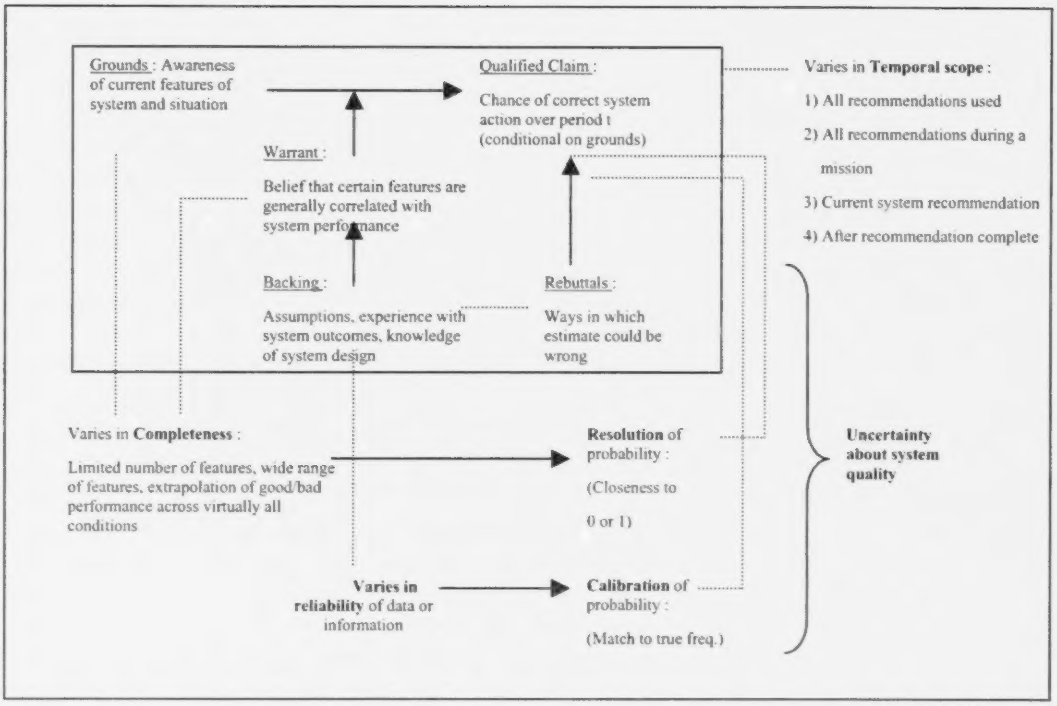

Grounds : Awareness of current features of system and situation
Qualified Claim :
Chance of correct system action over period t (conditional on grounds)
Warrant :
Belief that certain features are generally correlated with system performance
Backing :
Assumptions, experience with system outcomes, knowledge of system design
Rebuttals :
Ways in which estimate could be wrong
Varies in Temporal scope :
1) All recommendations used
2) All recommendations during a mission
3) Current system recommendation
4) After recommendation complete
Varies in Completeness :
Limited number of features, wide range of features, extrapolation of good/bad performance across virtually all conditions
Resolution of probability :
(Closeness to 0 or 1)
Uncertainty about system quality
Varies in reliability of data or information
Calibration of probability :
(Match to true freq.)

Figure 7: Cohen, Parasuraman & Freeman (1998) Argument-based Probabilistic Trust (APT) Model

In the APT model, the *qualified claim* represents an objective degree of trust in the system dependent on the features of the system (*grounds*), the belief that system features are correlated with performance (*warrant*), and the operator's background assumptions, experience and knowledge supporting these beliefs (*backing*). The outcome of the evaluation process represents the probability that a particular course of action is correct or, in other words, how much the human agent can trust the automated system output. These probability estimates may be incorrect in a number of ways (*rebuttals)*.

The other important aspect of this model is its ability to account for the changes in trust over time. First, this model accounts for the duration of the trust assessment in that it captures the relationship between different kinds of trust decisions and the *temporal scope* that is associated with it. Second, it reflects operators' varying levels of understanding of the factors affecting trust over time (*completeness*). Third, the model takes into consideration the degree to which an operator can reduce uncertainty about the decision aids (*resolution*). Fourth, the model also accounts for the quality or amount of information associated with the trust appraisal (*reliability*). Finally, similar to Muir (1994), the model considers the importance of assessing how well operators' trust corresponds to the true quality of the automated system (*calibration of trust*). Overall, Cohen et al., (1998) claim that "*the most important use of APT is to chart how trust varies, from one user to another, from one decision to another, from one situation to another, and across phases of decision aid use*".

A key strength of this work is the depiction of trust development as representing an *event tree*. An example of this is provided by Cohen (2000) describing Army aviators using a fictional version of a military decision aid.

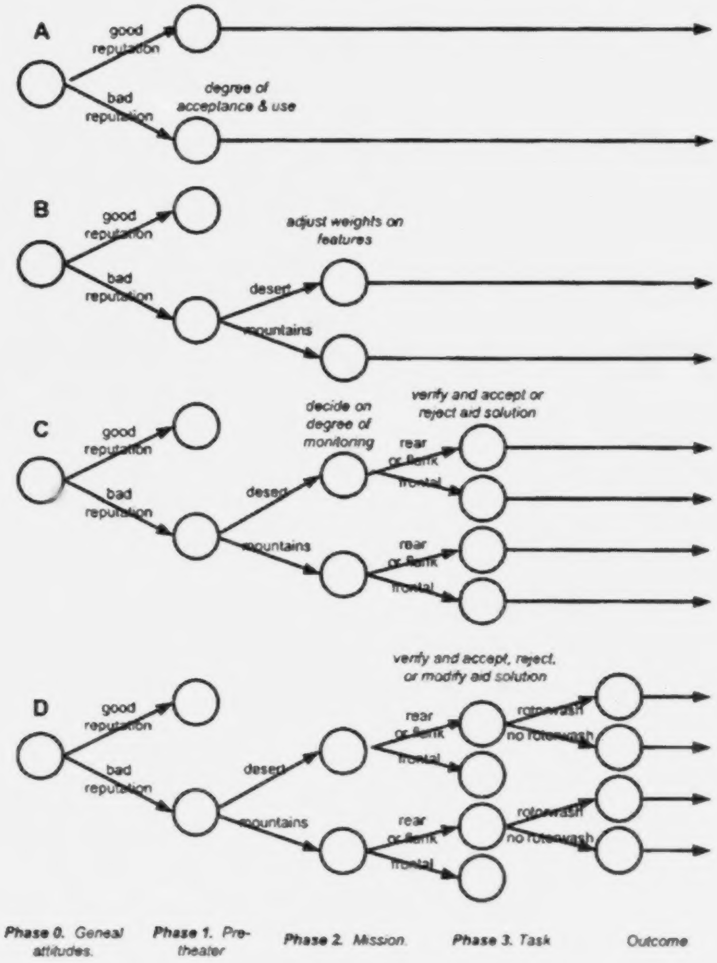

A
good reputation
bad reputation
degree of acceptance & use
B
good reputation
bad reputation
adjust weights on features
desert
mountains
C
good reputation
bad reputation
decide on degree of monitoring
desert
mountains
verify and accept or reject aid solution
rear or flank
frontal
rear or flank
frontal
D
good reputation
bad reputation
desert
mountains
verify and accept, reject, or modify aid solution
rear or flank
frontal
rear or flank
frontal
rotorwash
no rotorwash
rotorwash
no rotorwash
Phase 0. Geneal attitudes.
Phase 1. Pre-theater
Phase 2. Mission.
Phase 3. Task
Outcome

Figure 8: Event Tree (Cohen, 2000)

Factors that impact on trust development begin at the general attitudes stage, and continue through pre-theatre, missions and tasks, and eventual outcomes. This approach to understanding the trust development process, although very cognitive in its current implementation, is unique in its focus within the military domain. Consideration should be given as to how to incorporate the event tree approach into understanding trust in automation in future work.

6.5 Seong and Bisantz (2000)

Seong and Bisantz (2000) have created a model of human trust in automation based on the Lens model. The Lens model, developed by Brunswik in 1952, represents information transformation from stimulus (information presentation) to response (judgement), arguing that humans process information based on the observation of cues about the true state of the environment and then make an informed judgement, or response. Thus, the three important components of the Lens model are:

1. the true state of the environment
2. cues observed by an operator about the true state of the environment
3. the judgement by an operator based on the cues observed

Llinas, Bisantz, et al., (1998) assert that the Lens model is appropriate for modeling human trust as it examines both the cues used by the operator in forming trust and how well these cues reflect the true situation. The application of the Lens model to human trust in automated systems is conceptually straightforward. As seen in Figure 9, the level of trust represents the operator's judgement of trustworthiness, trustworthiness represents the actual trustworthiness of the automation, and observable characteristics represent the cues of the system's trustworthiness observed by the operator.

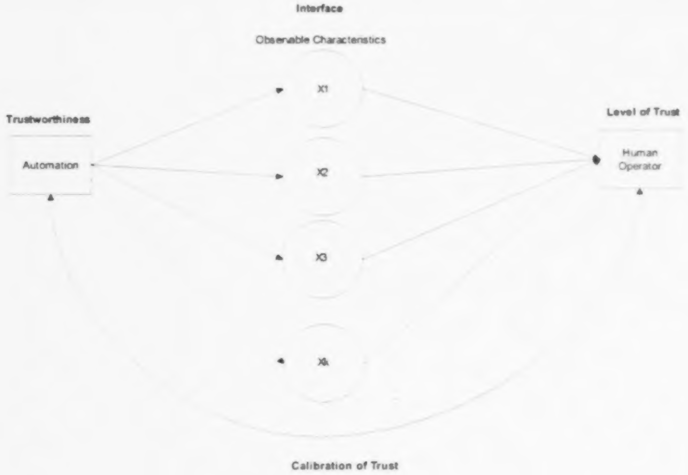


Figure 9: Model of human trust in automation using the Lens model (Seong & Bisantz, 2000).

Calibration of trust, shown at the bottom of the diagram, is achieved when the operator's assessment of trustworthiness matches the true state of the environment.

As outlined in Llinas, Bisantz, et al. (1998)[5], there are, however, some practical problems in applying the Lens model to human trust in automation. First, applying the model requires knowledge of true trustworthiness of the system. This, of course, cannot be measured. Second, the operator's level of trust can be measured, but only subjectively. In order to circumvent these problems, the authors argue that trust judgements should be based on performance rather than on subjective ratings. Then, an operator's utilization of a system component will reflect their trust in that component. Several cues such as predictability, dependability, reliability, competence or robustness will influence operator's judgements about whether to use automation. The authors assert that all of these factors can be measured objectively. The "true state" of automation, they argue, can be conceptualized as the adequacy of the controller, or as the integrity of the data source. By reconceptualizing key elements, then, the authors of this Lens model argue that trust in automation can be measured.

---

[5] This backdating occurs because the model was formalized only later in the project.

This model is admirable in that it includes the notion that an operator's decision to use a system depends upon his interpretation of characteristics of the system such as reliability and dependability. However, a major conceptual limitation, in our view, is that the model assumes that the operator's use of the system reflects their level of trust in it. As previously outlined, there are many factors that may influence one's decision to use automation, only one of which is trust. Perhaps more critically, however, even after reconceptualizing key elements, it is unclear how the adequacy of the controller, or the integrity of a data source can be accurately measured without extrapolation.

6.6 Madsen and Gregor (2000)

Madsen and Gregor's model of human computer trust (HCT), shown in Figure 10, is based on the notion that an operator's overall trust in an automated system comprises a cognition-based component and an affect-based component.

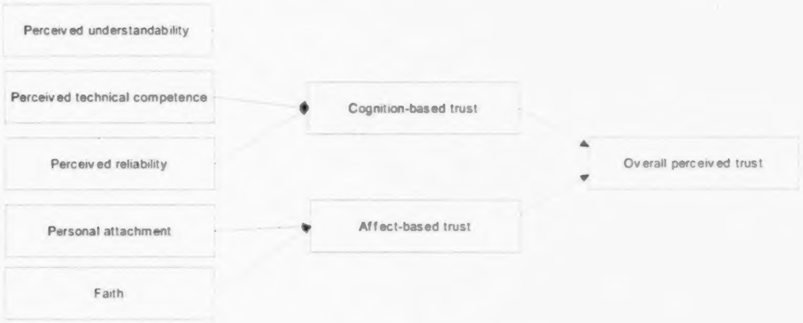


Figure 10: Model of Human-Computer Trust Components (Madsen & Gregor, 2000)

Madsen and Gregor (2000) define the cognition-based component of human-computer trust as that which is based on the operator's intellectual perceptions of the system's characteristics. Conversely, the affect-based component is based on the operator's emotional responses to the system. Consequently, the affect-based component plays a greater role in situations in which the operator does not have sufficient knowledge of the system upon which to base a cognitive decision.

The model created by Madsen and Gregor (2000) is valuable in that it makes a distinction between cognition-based and affect-based trust similar to the interpersonal trust literature. As well, the model includes most of the factors influencing trust that have been identified by other researchers, such as system reliability, competence and faith. Nevertheless, this model does not attempt to capture the process in which trust in automation develops, nor the dynamics of how trust changes as an operator gains experience with a system. In light of the problems of interpretation related to the measures of human/computer trust (see Chapter 9), we would argue that the model has not yet been validated at this point and more empirical research is needed.

6.7 Kelly et al., (2001)

Kelly et al. (2001) created another model of trust in automation that purports to represent all the factors influencing trust as well as the relationship between the factors. The model was developed as part of the process of creating a subjective measure of trust in automation specific to the air traffic control domain. These researchers also proposed that, as well as illustrating the relationship between the trust factors, the model could provide a framework for automation design principles and guidelines.

System competence, the operator's understanding of the system and the operator's self-confidence are three main factors influencing an operator's level of trust in an automated system, as depicted in Figure 11.

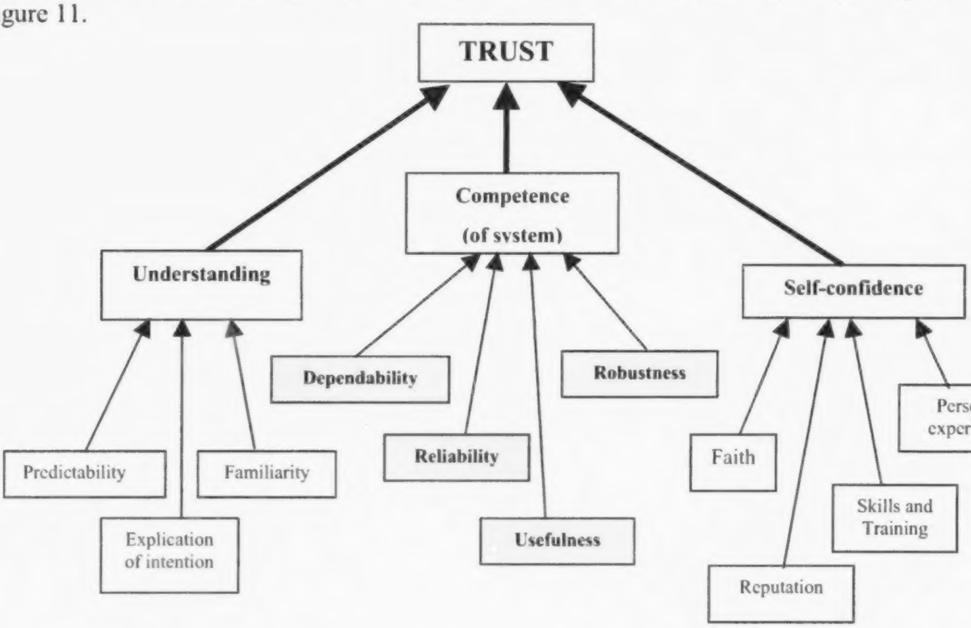

TRUST
Competence
(of system)
Understanding
Self-confidence
Dependability
Robustness
Personal experience
Predictability
Familiarity
Reliability
Faith
Skills and Training
Explication of intention
Usefulness
Reputation

Figure 11: Kelly et al., (2001) model of trust and the relationship between factors.

Each of these components is, in turn, influenced by several factors. This model makes some important distinctions. First, it distinguishes between the actual performance of the automation and the user's own understanding of the automation. This is an important distinction, in the sense that our views of the trustworthiness of automation are not absolute, but relative. The fact that automation performs reliably does not necessarily imply that it will be seen as trustworthy. If an automated system cannot be understood by operators, it may not be trusted regardless of its performance. To its credit, this model is one of the few to directly incorporate the notion of skills and training as influencing trust in automation. Showing skills and training to be related to self-confidence, however, suggests that skills and training only impact through the user's self-image. We would expect it to impact not only on self-confidence, but also directly during interactions with automation.

This work makes a number of important contributions to understanding how to measure trust in automation. First, it is based on a distributed cognition "Rich Picture" influence diagram showing the many influences on an air traffic controller's trust, as indicated in Figure 12:

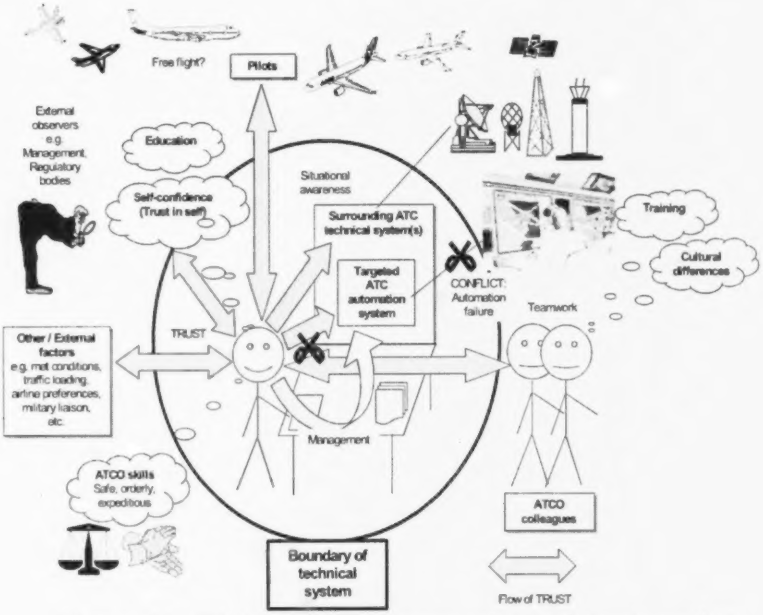

Free flight?
Pilots
External observers e.g. Management, Regulatory bodies
Education
Self-confidence (Trust in self)
Situational awareness
Surrounding ATC technical system(s)
Targeted ATC automation system
CONFLICT: Automation failure
Training
Cultural differences
Teamwork
TRUST
Other / External factors e.g. met conditions, traffic loading, airline preferences, military liaison, etc.
Management
ATCO skills Safe, orderly, expeditious
ATCO colleagues
Boundary of technical system
Flow of TRUST

Figure 12: Rich picture diagram of trust in automation within ATM context (Kelly et al., 2000).

This view of trust in automation considers both qualities of the trustor, in positing self-confidence to be an important aspect of trust in automation, as well as training, cultural differences and broader organizational factors (e.g. traffic loading). The inclusion of contextual factors is a critical improvement over other ways in which trust in automation has been viewed.

Secondly, this work also considers trust in automation at a broad level. The trust in automation component is one part of a larger effort aimed at understanding the entire world of an air traffic controller where situation awareness, teamwork and trust are argued to be the three principle components. Their relationship is depicted in Figure 13.

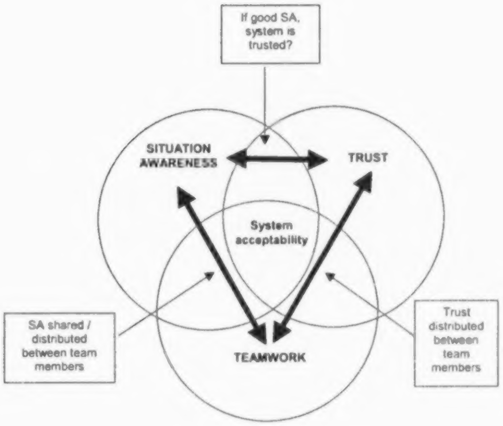

If good SA, system is trusted?
SITUATION AWARENESS
TRUST
System acceptability
SA shared / distributed between team members
TEAMWORK
Trust distributed between team members

Figure 13: Trust in ATM Context (Kelly et al., 2000).

Within this diagram, it is clear that trust in automation is affected by (and simultaneously affects) situation awareness and teamwork. This is an important point. As we noted in an earlier chapter, thinking about trust in automation in isolation of trust in the human operators of the automation (e.g. the other members of one's team) makes little sense.

On the other hand, this model has yet to be empirically validated. Although the work has been validated at an informal level, a report that speaks to more formal validation efforts was not available at the time of this review but, according to the project timelines, should be in progress. Further, the model is limited in that it only represents the factors influencing trust and does not attempt to describe the trust development process or the way in which trust changes as operator experience changes.

6.8 Conclusions

The models presented above are valuable and significantly contribute to the knowledge base of human trust in automation. The models of Kelly et al. (2001), Madsen and Gregor (2000), Seong and Bisantz (2000), Riley (1994) and Muir (1994) all consider several factors that influence human trust in automation.

One of the problems in the trust in automation literature (evident in trust literature generally) is that theorists often use different terms in order to represent the same (or very similar) construct. This has lead to a proliferation of terms with very similar meanings (e.g. complacency, non-vigilance, overtrust). Without careful examination of these terms and their underlying meaning, it may be very difficult to recognize the commonality in the key factors seen to influence trust in automation. This being said, it is important to note that there is relatively good agreement on the key factors that influence trust in automation, although factors related to the properties of automation have received the bulk of the attention.

In this sense, we would argue that focusing on the properties of a system to the exclusion of the properties of an operator leaves an important part of the story out. In general, operator self-confidence is the only factor has typically been considered by trust in automation models. Several other factors, such as trust history, overall propensity to trust etc. are also likely to influence decisions to trust automation. Moreover, most models also do not appear to seriously consider contextual factors as an influence on trust in automation. The extent to which trust in automation will develop is a product not just of the properties of the automation, or even of the operator. The development of trust in automation will also be influenced by the context in which trust decisions occur. All of these factors are described more fully in the next chapter.

In addition, these models do not account for the dynamic nature of an operator's trust in an automated system and generally do not describe the manner in which it develops over time. The time-series model by Lee & Moray (1992) accounts for the dynamics of trust but includes only fault characteristics and previous system performance as the variables influencing trust.

In general, the models discussed in this chapter have not been adequately validated. A few of the models have been validated empirically to some extent (Muir, 1994; Muir & Moray, 1996; Lee & Moray, 1992; Lee & Moray, 1994; Riley, 1994; Madsen & Gregor, 2000). Given that many of the models are quite complex, (e.g. Riley's model includes a total of 14 variables), empirical validation would be a tremendous undertaking. Nonetheless, some validation efforts have purported to show that, in at least some cases (e.g. Lee and Moray, 1994), trust in automation can be predicted very well from the actual physical properties of the automation (Kelly et al., 2001). If automation performs reliably, it will be trusted. It is critical that this conclusion is also tempered by attentiveness to the limited context in which this research was performed. In interactions with more complicated systems, there may be more room for other variables to exert influence, in the sense that more complicated forms of behaviour are possible. As such, the generalizability of this research should be questioned.

From a similar perspective, it is also important to note that the conclusions that can be drawn from this kind of research are perhaps limited by the fact that there is clearly the potential for demand characteristics to extent an influence. In a relatively simplistic "microworld", one might expect a substantial relationship between the properties of the system, and trust in the system. In situations where use or non-use of automation are the only possible behaviours, for example, it seems rational for participants to not use automation after a fault occurs or to continue to use it when it is working reliably. Similarly, when asked "*how much do you trust the system?*", it would seem irrational not to adjust one's trust downward if a fault had just occurred. The potential problem, however, is that given a wider range of possible responses, participants' behaviour in these experiments may be more complex and less predictable. In this sense, future research studying trust in such systems should focus more on possible demand characteristics, and should allow a broader range of behaviour.

Lastly, as important as these theoretical models have been in beginning to understand trust in automation, they have not been developed in order to understand trust in automation within the military context. Therefore what is needed is a fundamental model of trust in automation that includes the most influential factors yet also depicts the trust development process and how an operator's trust changes over time. Although the APT model by Cohen et al. (1998) was developed in the context of military tactical decision making, it does not appear to have been validated in a military setting or with any other types of automation that may be used in the military. The next two sections explore the factors that influence trust in automation in more detail, followed by a preliminary model of trust in automation.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 7 – FACTORS AFFECTING TRUST IN AUTOMATION

Many factors are likely to influence trust in automation. Some have been proposed at a theoretical level; others have been researched. The factors most likely to influence trust in automation are the properties of the automated system, characteristics of the user, and properties of the environment. These sets of factors are considered in the sections that follow.

7.1 Properties of the Automated System

A number of variables related to the properties of automation are likely to influence trust in automation.

7.1.1 System Reliability

System reliability refers to the extent to which a system does the job that it was designed to do. Automated systems that perform consistently well are more likely to be trusted (Kelly et al., 2001), than those that do not. Sheridan (1988) has argued that reliable automation (1988) "conditions" trust, in the sense that people come to generalize their expectancies from past experiences. This attribute is essentially the same as the dimension of *competence* (Muir, 1989) or *robustness* (Sheridan, 1988) previously discussed. In short, competent systems are likely to be trusted more than incompetent ones.

There is a good deal of empirical evidence showing that system reliability and trust are related (e.g. Moray & Inagaki, 1999; Riley, 1994). Work by Moray, Inagaki and Itoh (2000) argues that reliability is highly correlated with trust given by operators. In a study conducted in a simulated microworld, operators exerted either manual or automatic control over the system, in order to control the water temperature, and respond to system problems. Automation reliability was designed to vary as a product of the percentage of trials in which the automation suggested correct diagnosis and response to problems occurring during the simulation. After each trial, participants rated their subjective trust in the automation. Table 7 shows trust in the automation when functioning in automated and manual modes, as a function of varying levels of reliability.

Table 7. Subjective trust as a function of reliability and mode of control.

	Reliability					
	70%		90%		100%	
	Mean	StDev	Mean	StDev	Mean	StDev
Adaptive automation	4.8	1.3	7.6	1.3	8.4	.9
Manual control	5.8	1.3	7.5	1.0	8.6	.9

As the reliability of the automated system declined, so did trust, with trust varying falling from 8 to 5 on a 10-point scale. Clearly, declining system reliability can lead to a systematic decline in trust expectations. Just as importantly, these changes can be measured over time. In general, then, there is a good amount of evidence that the reliability of an automated system is a strong predictor of trust in the system.

There is also some evidence that only the most recent interactions with automation impact on trust perceptions. Research by Lee and Moray (1992), noted by Kelly et al. (2001), argues that people have a certain baseline of trust in automation situations, and this is affected primarily by the last one or two interactions. It is important to note, though, that this view seems overly optimistic in the sense that the impact of previous interactions with automation is surely affected by the outcome of the interactions. Very negative effects resulting from misplaced trust in automation are likely to exert a longer term impact.

7.1.2 System Faults

System faults are highly correlated with system reliability, but refer specifically to discrete events within the system. A good deal of empirical work has explored the relationship between system faults and the resulting impact on trust in automation. This work suggests that several different aspects of faults influence the relationship between trust and automation. In general, system faults do have a negative impact on trust in an automated system. During an experiment with continual system faults, trust in the automation reached its lowest point after only 6 trials, but trust did recover gradually even as faults continued (Lee & Moray, 1992). This work is consistent with the general observation in the literature suggesting that when faults occur during interactions with an automated system, trust is often quickly affected.

A consistent theme in the research is that the impact of faults on trust can be distinct from their impact on performance. There is some evidence suggesting, first, that trust is *less* resilient than performance in response to faults. After faults have occurred, for example, performance bounces back relatively well, but trust does not. Trust recovers much slower than performance, and even then does not recover to previous levels (Lee & Moray, 1992). This is consistent with other evidence suggesting that trust both disintegrates and recovers more slowly from system faults that occur (Moray et al., 1995); once broken, trust is somewhat less resilient. On the other hand, maintaining performance may be easier than maintaining trust, as a wider range of adaptive strategies can be used by participants in order to maintain their performance.

On the other hand, there is also evidence that trust is *more* resilient to faults than is performance. For example, in a study by Lee and Moray (1992), although fault did impact on performance (causing a 10% reduction in performance) these faults were slower to impact on trust (Lee & Moray, 1992). In short, the research suggests that trust does appear to be calibrated to the reliability of an automated system. But trust is also lost more rapidly when systems fail than it is regained when they perform reliably (Kelly et al., 2001).

Magnitude of Faults – There is also some evidence in the literature that trust in automation is also affected by the magnitude of a fault. In work by Lee and Moray (1992), different sized faults were studied, with the % difference between the actual and the target pump rate being the measure of fault magnitude. Results showed that relatively smaller faults (15% to 20%) had minimal impact on trust. Large faults (30% to 35%), on the other hand, caused trust to diminish, and to recover slowly. This effect on trust lasted for several subsequent trials. Performance, on the other hand, was not affected by the fault. This suggests that the magnitude of a system's fault can affect trust independently of affecting performance (Lee & Moray, 1992).

Variability of Faults - Although, in general, large faults have a more negative influence on trust development than small faults, other evidence suggests that the variability in fault presentation is perhaps even more critical than the size of the fault per se. In work by Muir and Moray (1996), for example, small faults of varying magnitudes diminished trust more than large constant errors. As such, fault variability also influences trust in automation.

Locus of Faults - When faults occur in a specific part of an automated system, a key issue is the extent to which trust in the system as a whole -- versus in its constituent parts -- will be affected. There is empirical evidence that the impact of a fault sometimes spreads and sometimes does not.

Seminal work by Muir and Moray (1996) found that the distrust did spread between components of a system. Distrust in a poorly functioning pump display affected levels of trust in the pump's (competent) control system. This suggests that when trust or distrust in automation spreads, it might also generalize to other functions controlled by the same subsystem. This may lead to unwarranted distrust. It is important to note, however, that the distrust did not spread to independent but similar systems.

Advance Knowledge about Faults - Empirical research shows that when people have prior knowledge of faults, these faults do not necessarily diminish trust (Riley, 1994). This is a fascinating finding, and one that is worthy of further study. One of the reasons that this might be the case is that advance knowledge may alter the dimension of risk. Given information about the fault, people still know that the automation will fail, but the very predictability of the failure may be comforting. As such, there is some reason to argue that, at least in the automation domain, predictability can be important than absolute reliability.

Overview of Faults - In general, findings related to the impact of faults on trust are less than conclusive. Faults do impact on trust, and trust can be quickly broken. On the other hand, trust in automation can be restored with fault-free performance, but recovery to initial levels of trust can take a long time and may never return completely. The literature suggests that the scope of the impact depends on specific qualities of the fault (e.g. magnitude, etc.). Trust is generally more affected by large faults, and can be relatively unaffected by small faults. A caveat is that the variability of faults may be more critical than the size of the faults. Trust can be both less resilient to system faults than performance and more resilient than performance.

7.1.3 System Components

Operators' level of trust in an automated system is not always a gestalt. Trust in automation can also be a product of expectations about the constitutive elements of the system. These elements are:

1) Display format - the graphical representation, the design, the logical / dispositional layout, and the accessibility of the information provided by the automated device;

2) Hardware or control system - comprised of the physical building blocks that constitute the automation (e.g., engine, computer, quadrants, joysticks, levers, etc.) and of the associated mechanistic and electronic control apparatus;

3) Algorithm (software) - a set of steps or procedures used to combine, synthesize, and analysed the information present in the environment.

Display format and the hardware or control system of automation can both be seen as having distinct qualities likely to influence trust. Both can be observed, and both can give either implicit or explicit information about the probable trustworthiness of the system. Displays that provide clear and understandable information about the status of the automation, for example, are likely to elicit more trust. Similarly, hardware that is well crafted and rugged also provides a strong basis for positive trust expectations.

The underlying algorithms of automation are less observable. The algorithms underlying automation can affect trust from several different perspectives. First, faulty algorithms are more likely to result in inferior performance. Also critical, however, is the "accessibility" of the algorithm to the automation operator. Accessibility refers to the extent to which a user can access the decision rules used by the automation, and the extent to which these processes are both literally and figuratively comprehensible to the average user. Even the best algorithms will not necessarily promote trust unless they mimic the reasoning and decision making processes commonly used by humans (Simpson & Brander, 1995; in Aldern, 1995). Development of trust requires predictability and a belief that the automation is correctly prioritizing and handling information.

At such, in order to fully understand trust in automation, one must consider both the whole (i.e. the system), as well as the parts that make up the system.

7.1.4 System Transparency

The transparency of a system is also likely to influence the extent to which the system is seen to be trustworthy. Transparency, of course, refers to the extent to which an object is "clear or easily discerned" (Oxford Dictionary). By definition, automation that is more easily discerned will be more transparent, and automation that is more transparent will be more trusted than automation that is not transparent.

In the available literature, several concepts that closely parallel system transparency are noted; the most prominent is explication of intention (Sheridan, 1988). The term "explication of intention" seems to imply that automation is specifically charged with responsibility to give information about its inner workings. Thinking about automation as having intention may make sense in thinking about some forms of automation, but seems to have less meaning when thinking about simpler forms of automation. The concept of system transparency, however, does not require the assumption that automation is in some way motivated to give or withhold information. It merely refers to the extent to which the inner workings of the system are accessible and available to the average user.

Automation can be transparent in several ways, in terms of both the process by which automated systems perform, and in the product rendered by the automation. In terms of process, systems with more transparent explanation facilities will be more likely to be trusted, as these facilities make system functioning more easily understood (Simpson & Brander, 1995; in Aldern, 1995). Automation with more complex self-explanation abilities may have an advantage when trust violations occur. Such automation can explain the processes that it is undertaking, as well as to give information about why it is not performing its assigned task. Automation having an enhanced ability to explain itself, then, may assist in buffering the impact of trust violations. Other authors have argued that the ability of automation to "explain itself" may allow the operator to query the system during periods of low activity (Simpson and Brander, 1995; in Alder, 1995), and to build trust incrementally. This allows for trust to develop during "down" times, in preparation for trust to play a role at more critical points. The extent to which automation is transparent in explaining what it is doing and why, then, will influence its ability to garner operator trust. Indeed, one of the ways that this might occur is that transparent systems will be more likely to aid operators in forming a mental model of the system.

The transparency of automation output is also likely to vary. It is possible to imagine considerable differences in the discreteness of an automated system's output. In cases where an automated system functions with a less tangible final product, for example, taking a snapshot of its performance, and being able to evaluate this outcome against the desired outcome may be very difficult. For process control automation, the output is often the physical product, and it is relatively easy to access the quality of this product. In the case of automation that provides information, the product is the recommendation made by the decision aid. It may be much more difficult to accurately access the quality of information than to access the quality of a paper product. In the longer term, this may lead to a failure to calibrate trust effectively. All of these factors related to the transparency of system process and performance are likely to influence trust in automation.

7.1.5 Level of Automation

The level of automation that a system reaches (that is, the relative allocation of function to a human or to the automation) will influence the development of trust in automation. The development of trust in automation may be negatively correlated with levels of automation – that is, it may be more difficult to trust higher levels of automation. This is true from a number of perspectives. First, it has been suggested by Muir (1989; 1994) that an important factor for trust development is the understanding of the automated system functioning. As automation takes over more functions, it may get more difficult for operators to understand the system, and to evaluate its performance.

Limited empirical work has studied the dynamics of trust across the different levels of automation, but there is some reason to believe that systems with varying levels of automation have different implications for trust. Empirical work based on a relatively simple level of automation (e.g. SVL 3) exploring the relationship between trust, self-confidence and the use of automation did not obtain the same results when attempted within a more complex simulation environment (e.g. SVL 7) (Moray, Inagaki and Itoh, 2000). This work suggests that the relationship between trust and the use of automation may not always hold with more systems that function at a higher level of automation. Future studies are required to better understand the impact of varying the level of automation on trust in the automation.

7.1.6 Interactivity of the System

Automated systems that provide more opportunity for interaction between the human and operator, on average, would seem more likely to be trusted. This speculation is based on the fact that trust is seen to develop over time, as the result of increasingly elaborated predictions and expectations. The more that one interacts with automation, the more that these predictions and expectations are developed

---

(assuming, of course, that the automation performs reliably). The importance of interactivity in working with automation is expressed in principles of user-centered design.

7.1.7 Susceptibility to Tampering

Llinas, Bisantz et al. (1998) have pointed out that the susceptibility of a military system to tampering is important for the establishment of trust, especially in an information warfare context. Within this context, the recommendations that decision aids provide can be understood as having been intentionally degraded (e.g. sabotage), intentionally with camouflage (subterfuge), or unintentionally. Within the military context, then, the extent to which automation is seen as susceptible to tampering from external parties will influence the trust in automation that can develop.

7.1.8 Predictability and Dependability

The predictability and dependability of automation will also impact on trust. Automation that can be predicted, either through reliable performance, or through consistent (even flawed) performance is more likely to be trusted. This consistency in behaviour enables one to form broader attributions about the trustworthiness of the automation at a broader level. As noted earlier, predictability and dependability are both noted in previous research and theory to influence trust (e.g. Muir, 1994).

7.1.9 Reputation of System Designer

The designer of an automated system (e.g., company, etc.) can affect trust in automation. Muir (1994) cites knowledge about the maker of an automated system as impacting on trust development. To the extent that a system designer has a known reputation, the reputation may guide expectations about the trustworthiness of the automation. Moreover, knowledge about the designer of a system can also provide information about the standards used to make the system, and even about the values and priorities that the designer is likely to have incorporated . These forms of reputation may impact on trust in the automation especially in the early stages of use. As a person gains experience with the automation, direct experience in terms of system reliability etc. is likely to assume a dominant role.

7.1.10 System Appearance

Although not prominently noted in the available trust in automation literature, the very appearance of automation may be a predictor of trust in automation. The apparent quality of the system, the ruggedness of its design, may all provide important clues as to its trustworthiness, and perhaps its reliability over the long term.

7.2 Properties of the Trustor

Properties of the human who interacts with the automation are likely to influence trust in automation.

7.2.1 Propensity to Trust Automation

In the context of social learning theory, Rotter (1967) has suggested that people differ in terms of their predisposition to trust others. This predisposition is conceptualised as a generalized tendency to trust others and is hypothesized to be a relatively stable personality characteristic. In the context of the present review, it is suggested that a general tendency to trust automation is likely to affect trust in specific forms of automation.

Propensity to trust automation not received much empirical attention in the literature. There is some evidence, however, that operators' overall propensity to trust automation is distinct from trust toward a specific automated system (Parasuraman & Riley, 1997). It is possible for an operator to have a very positive attitude toward automation in general, but to have a very low level of trust in a specific automated system.

7.2.2 Ability to Form Mental Model of System

People naturally vary in their ability to form mental models of automated systems. This ability may involve knowledge of the inner workings of the automation, and the development of coherent knowledge systems and expectations about the automation. As trust rests partly on being able to understand and to develop positive expectations about such systems, people who are unable to develop adequate mental models are also likely to have more trouble trusting automation. With inadequate models, it is very difficult to be able to predict future system behaviour.

7.2.3 Trust History

The outcome of previous trust relationships with automated systems (i.e., trust history) is also likely to affect trust in automation. The development of trust in automation has been described as an "ascending function of experience, fostered by a perception of consistent and desirable behaviour" (Muir, 1994). The prior interactions of an operator with automation are thus likely to give rise to expectations, beliefs and feelings about the reliability of the automation, and this trust history will affect trust in the automation over time.

7.2.4 Self-Confidence

Self-confidence is an individual difference factors that has been extensively studied in the trust in automation literature. Early work suggested that the relationship between trust and self-confidence was a relatively simple one (Lee & Moray, 1994). When trust exceeded confidence, automation would be used rather than manual control. But, when self-confidence was higher than trust, manual control would be used. Work by Riley (1996) also looked at the role of trust and self-confidence. In four separate experiments, Riley showed that trust and self-confidence are important determinants of automation use.

Subsequent work investigating more complex systems, however, suggested that this simple relationship between trust, self-confidence and the use of automation does not always hold (Moray, Inagaki, & Itoh, 2000). Whereas the earlier work was based on a relatively simple level of automation (e.g. SVL 3), work conducted using a more complex simulation environment (e.g. SVL 7) did not obtain the same results. Time series modelling was used to explore the factors bearing on these new findings. In general, trust was shown to be influenced by properties of the system (e.g. real or apparent false diagnoses), whereas self-confidence was affected by the experiences of the operators (for example, whether they had been responsible for accidents). Moreover, self-confidence was not affected by system reliability (Moray, Inagaki & Itoh, 2000).

Work by Lewandowsky et al. (2000) also suggests that self-confidence need not be adversely affected by shifts in the reliability of automation. Furthermore, this work also argues that the relationship between trust, self-confidence and the reliance on a target such as automation is more powerful in the automation domain than in the interpersonal one.

7.2.5 Personal Work Style

The majority of research on trust in automation has been limited to microworlds. In these kinds of simulations, the scope of an individual's interaction with automation is specifically delineated. Many other interactions with automation, however, allow more personalized interaction. The extent to which there is a match between how operators naturally work and the "work style" of automation will also impact on trust in automation. High levels of congruence should make it easier for trust to develop, as high overlap is likely to make one better able to predict how automated systems will perform. The impact of personal work style on trust in automation has been all but overlooked in the existing literature. In theory, if it were possible to identify the factors most diagnostic of trustworthiness within the work domain for a given person, it may then be possible to tailor the automation's performance (and hence perceived trustworthiness) to these dimensions.

7.2.6 Age

There is clear agreement at both the theoretical and anecdotal level that age is likely to affect peoples' ability to trust automation. Older people typically have less experience using automation, and may not have developed the strategies and models that they need to understand (and hence to trust) it. It is noteworthy in this regard, however, that a literature survey conducted by Modalfsky and Kwon (1994) obtained inconclusive results concerning the influence of age on computer usage or attitude toward computers. Further research should be conducted to determine the relationship between trust in automation and age.

7.2.7 Cultural Influences

Cultural influences may impact on trust in automation: the way that technology is regarded in a culture (positively, negatively or neutrally) may influence how the people in that culture use it and trust it. In the Western world, wide range of experience and involvement with many forms of automation puts us in a better position to trust automation than would be the case in cultures without such opportunities.

7.3 Properties of the Environment

Several contextual factors are also likely to influence the trust relationship between humans and automation.

7.3.1 Risk

The general literature on trust identifies risk as a factor that influences the need to trust. Issues of trust are more likely to arise in high risk than in low risk situations, as trust is a critical means by which to reduce uncertainty. Risk may be one of the most important contextual factors influencing trust in automation.

Within the trust in automation literature, risk has been defined as "the likelihood and consequences of error" (Riley, 1996). Research exploring risk and the use of automation suggests that reliance on automation is moderated by the risk associated with the decision to use an automated system (Riley, 1996). In short, people are reluctant to use automation when the probability of adverse consequences is high (Riley, 1994, 1996). Similarly, it also takes people longer to reengage automation in high risk than low risk situations (Riley, 1994). Moreover, advance knowledge about the system behaviour may alter the dimension of risk. When people know when and how long the automation will fail, their trust is unimpaired and reliance on the automation continues to be high (Riley, 1996).

Work by Lewandowsky et al. (2000) sought to manipulate risk in order to understand the impact on trust in automation. This was done by altering the speed at which an automated system ran (i.e. the overall throughput of the system) in some of trials. Results showed that, contrary to prediction, ratings of trust in the system were significantly higher in the high speed trials than in the low speed trials. But since the high speed trials always followed the low speed trials, this finding is attributed to a build-up of trust during prolonged experience with the automation. The manipulation of risk in this experiment was somewhat problematic in the sense that participants' actual perception of risk was not measured. Nonetheless, it is clear that higher levels of risk will typically necessitate higher levels of trust in automation. As people are more hesitant to use automation when risk is high, then even higher levels of trust will be needed in high risk situations.

7.3.2 Operational Context

Operational contexts differ in terms of the factors that may influence trust in automation – for example, physical stressors, risk, time pressure, etc. The levels of these factors will typically be higher during actual operations than in garrison.

There are typically more opportunities to use automation during an operational deployment than during exercises or simulation. Based on our studies of interpersonal trust in small teams (Adams and Webb, 2003), we might expect that trust in automation when in garrison, for example, may well have a somewhat different quality than trust in automation in actual operational contexts. Tank crews argue that trust in other teammates has a somewhat conditional, provisional quality. Crew members asserted that it was possible to trust other teammates, but that this trust could not be fully developed in the absence of shared operational experiences. The same is likely to be true of trust in automation. A user may form expectations about how trustworthy automation is likely to be under less demanding conditions, but these expectations are fully tested only when the automation is functioning in adverse conditions. Only then would fully developed trust in automation be possible.

7.3.3 Organizational Factors

Several factors related to an organization are likely to influence trust in automation. For example, because of the need for accountability within an organization, policies regarding the use of automation may be put in place. If they are restrictive, these policies have the potential to signal uncertainty about the automation, and to make operators wary and less likely to trust it. On the other hand, having a broader set of rules governing the use of automation could also improve trust in automation, since these rules may ultimately make the automation more predictable.

It is important to note that stricter policies on the use of automation may not be a reflection of lack of confidence on the part of the military. As Miller (1995; in Aldern 1995) pointed out in the keynote address to the Human/Electronic Crew symposium,

> "..the rules do not reflect the Commander's confidence in his weapon systems. They are also the politician's final means of controlling events on the battlefield. Thus, even if we develop intelligent, predictable and trustworthy systems, this progress is more likely to translate into an increased confidence of success, rather than greater freedom of operation, and we cannot expect a sudden change in the way we go about our business."

In essence, this suggests that the use of complex systems and ultimately the trust that can develop in them may be constrained by other factors like the political and societal context in which the automation resides. As such, within the military system, trust in automation product of trust within the system as a whole, not just in the actual equipment.

In this sense, an organization's history and policies regarding automation are also likely to be relevant. There are several notable examples of faulty automation that have not been dealt with by the political and/or military system. With Canadian Forces, for example, problems with the Sea King helicopter have been very well publicized (e.g. Northup, 2002). Operators may be less likely to trust automation, a priori, if there is a prior track record of problematic automation being brought into service. Although not likely to exert a huge impact, organizational history could impact negatively on the development of trust in automation.

Lastly, it is important to remember that automated systems also stand within a specific interpersonal context. We argued earlier that the interpersonal context in which automation is embedded may also impact on trust in automation. Within military teams with low trust, for example, trust in automation may also be compromised, as the very use of automation often requires skill with the automation and the ability to interpret the outcomes that automation provides. Similarly, soldiers who feel unsupported by their superiors may be inherently less likely to trust automation because it is seen as an extension of a system that may not be worthy of their trust. In this sense, it seems important not to completely separate trust in automation from the larger issue of interpersonal trust.

7.3.4 Training

The level of training that prospective operators receive before interacting with automation has the potential to influence the development of trust. Since trust is based on being able to make confident predictions about how the automation is likely to behave in a given situation, higher quality training should, in theory, promote more trust in the automation. Put another way, if the automation is better understood, it should be better trusted. Another factor that may influence trust is the degree of experience with the automation.

It is important the operators be trained to understand the limitations of automation, as well as its capabilities, although as previously noted, limitations need not impact negatively on trust. As such, training needs to be designed to reflect automation predictability in terms of both desirable and undesirable performance.

7.3.5 Task Demands

The extent to which a given piece of automation can be trusted may also depend on the actual task that it is being asked to do. Tasks vary in terms of their complexity (e.g. the number of steps required to complete the task), and their difficulty. With tasks that are more demanding, there is more room for error. We would therefore expect that automation would be less trusted when performing highly difficult tasks than when performing less difficult tasks. There is no empirical evidence that speaks to this issue in the literature, although there is some evidence in related areas. Research has shown that task difficulty does influence the use of automation. McFadden, Giesbrecht, & Gula (1998), for example, have shown that operators continuously used an automatic tracker regardless of its decreased reliability when their task was more difficult. Although this finding does not specifically address trust, there is some reason to believe that the same may also be true of trust. However, the use of automation in this study was not volitional, but in fact, a product of workload. In this situation, the operator may be hard pressed to switch from automatic to manual control, due to workload. As such, in thinking about the relationship between trust and workload, it is critical to consider whether the use of automation is volitional or not.

7.4 Research Implications

In summary, several different sets of factors related to properties of automation itself, properties of the trustee, of the interaction and of the context in which decisions to trust automation occur are likely to influence the development of trust in automation. The following issues emerge as important considerations for future research.

1. The trustworthiness of automation is likely to be affected by many factors. In creating measures of trust in automation, it will be important to understand trust in the subcomponents of the automation (e.g. display, controls), as well as trust in the automation as a unit.
2. Particularly within the military domain, high levels of risk and uncertainty inherent in many situations make the need to understand trust in automation even more critical. In the context of a research program, the extent to which trust in automation is affected by contextual factors such as risk and uncertainty needs to be a critical focus. At an experimental level, manipulating risk and uncertainty will aid in understanding changes in trust as a psychological state.
3. One of the factors that influences people's interpretation of faults is advance knowledge about the fault. More broadly, this finding would seem to argue that operators' trust in automation can perhaps be improved by ensuring that they know the true reliability of the automation, where the problems might be and when they are likely to occur.
4. The fact that trust is lost more rapidly when systems fail than it is regained when they perform reliably is an issue to be explored in future research. Are there system properties that predict this result? Does it hold for all forms of automation?
5. For future work exploring trust in automation and risk, it will be important to develop a different definition of risk, as existing definitions are problematic (e.g. risk is "the likelihood and consequences of error" (Riley, 1996). In our view, risk should be defined not only in objective terms (i.e. the actual level of risk in a situation), but also in terms of risk actually experienced by an individual (i.e. perceived risk).
6. Within military contexts, trust in automation is likely to be influenced by the extent to which an automated system is open to potential tampering, the extent to which hostiles may degrade the quality of both functioning and output within adversarial environments.

CHAPTER 8 – CONSEQUENCES OF TRUST IN AUTOMATION

Much of the interest in the topic of trust in automation relates to the assumption that operators who trust automation will accrue some benefit from its use, either in terms of support for task processes or task output. Despite the seemingly obvious nature of this assumption, however, it is important to note that it has yet to be adequately tested at an empirical level. The following sections review the literature related to the consequences of trust.

8.1 Problems with Terminology

Understanding the consequences of trust in automation has been problematic, in part, because of conceptual confusion in the literature. Terms such as undertrust and overtrust, for example, are frequently used in the literature, but seem to be used inconsistently. Some researchers assert that overtrust and undertrust are forms of *mistrust*, which is different from *distrust* (as discussed below). Parasuraman and Riley (1997), use the term *misuse*, rather than mistrust, to refer to both overtrust and undertrust. Disuse is different from misuse and is defined as the case in which automation has been inappropriately allocated to a certain type of task, when that task is more suitably carried out by a human.

Similarly, the term distrust has been used in several ways in the literature. For example, Llinas, Bisantz, et al. (1998) define distrust as automation operating beyond its judged boundaries of competence. Mistrust, on the other hand, is defined as the situation in which a human has an incorrect level of trust in the automation (e.g. trusts an incompetent system too much or doesn't trust a competent system). Other researchers use mistrust and distrust interchangeably, suggesting that complacency is an example of distrust as well as mistrust. Still others have defined distrust as appropriate from a normative perspective, and as occurring when an operator does not trust an unreliable automated system (Kelly et al., 2001).

In order for trust in automation research and theory to progress, it will be important to clarify the meaning of these many concepts. We argue that some terms should be eliminated, as they fail to make a substantive contribution to describing the true phenomenon of interest. The terms, undertrust or overtrust, for example, are often used to describe the behaviour of failing to use automation enough, or using it too much. From our perspective, it is important to disentangle trust as a psychological state from trust behaviour as much as possible. Only the behavioural *correlates* of undertrust and overtrust can be observed. As such, rather than resorting to concepts that imply the existence of trust as a psychological state, it makes more sense to simply describe the behaviour that is occurring (e.g. using automation or not using automation) without extrapolation to the concurrent psychological state.

8.2 Trust and Use of Automation / Reliance on Automation

In the military context, there are many situations in which automation must be used to ensure optimal performance. Trust in automation has been widely argued to influence (and to even determine) whether or not automation is used. This section considers the empirical evidence linking trust in automation and reliance on automation.

Muir (1994) and Muir & Moray (1996) found that an operator's trust in an automated system was positively correlated with the percentage of time the system was used. Likewise, Muir & Moray (1996) found that as operator trust increased, use of automation increased. As trust diminishes, however, use of automation and reliance on automation is also likely to decrease. This conclusion is consistent throughout the reviewed research (e.g. Lee & Moray, 1992).

An important mediating variable affecting the relationship between trust and use of automation is self-confidence. Lee and Moray (1992) found a positive – albeit low – correlation between trust and use of automation. This argues that trust alone does not guide the percentage of time spent using automatic control. Based on this, Lee and Moray argued that trust, coupled with self-confidence, might provide a better explanation for operator's choice of manual or automatic control. Subsequent work by Lee and Moray (1994) showed a strong relationship between the difference in operators' trust and self-confidence and their reliance on automation. Lewandowsky et al. (2000) replicated this finding and showed that the difference between trust and self-confidence is a good predictor of automation use. As a whole, then, this research suggests that there is a strong relationship between trust and the use of automation, but is one that depends to some extent on self-confidence.

The majority of the research related to the consequences of trust, however, addresses trust in process control environments rather than trust in the information provided by decision aids. Very little research that we accessed during the course of this review directly addressed the important issue of trust in advice provided, for example, from an expert system. There is some empirical evidence that suggests that agreement with the information that an expert system provides may lead to less critical evaluation of the information. Work by Dijkstra (1999), for example, studied people's reaction to the advice of a decision aid provided about 3 law cases. In general, users who agreed with incorrect expert system advice provided about the legal cases required less mental effort, and spent very little time studying the advice. These participants also evaluated the legal cases as being easier than did participants who disagreed. On the other hand, these participants also scored lower on recall questions about the advice. This research addresses agreement with decision aids rather than trust in them, but the same may be true for trust in decision aids. Belief in expert systems may lead to less critical evaluation of their information because the systems are seen to be *expert*. The fact that trust in automation may lower the level of attentiveness paid to the output of a decision aid, or make operators less vigilant monitoring of the automation has the potential to be very problematic. This is particularly true within the military domain, where the cost of misplaced trust can be much higher. To the extent that automation is rarely – if ever – fool-proof, less critical evaluation of it may be problematic.

In the relationship between trust in automation and the use of automation, trust theorists have also argued that trust in automation is not unconditionally good. It is also important to consider the appropriateness of an operator's level of trust in relation to the reliability of the system. This issue is explored in Figure 14 below.

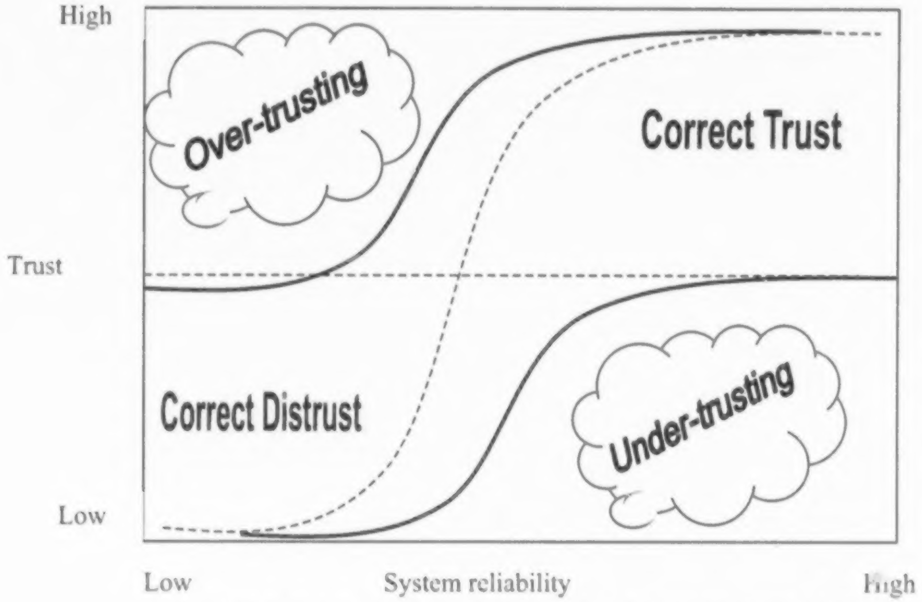


Figure 14: Theoretical relationship between trust in automation and system reliability (Kelly et al., 2001)

When a system shows low reliability, the correct response is to be distrustful of the system. Using the system too much in this diagram is called overtrusting. When a system is being reliable, it is correct to trust the system, and to allow it to exert control. This is obviously correct trust. Failing to use the automated system when it is actually reliable is also not optimal (and is called undertrust in this diagram, whereas not using unreliable automation is the correct response. As such, trust is not a simple uni-dimensional variable and the level of trust that is appropriate at any given time depends on the reliability of the system (Kelly et al., 2001).

The dangers of not having enough trust in automation have been explored at a theoretical level in the literature. When operators do not place enough trust in an automated system (relative to its reliability), they may resort to manual control and/or may ignore reliable alarms or safety devices. This can occur for a number of reasons, including inaccurate mental models, actual or perceived automation failure, or improper decision criteria. For example, if a decision criterion for sounding the alarm to warn an operator of a potential danger is too lax, false alarms may happen too frequently. This may reduce the operator's level of trust in the alarm system and the operator may subsequently turn it off. Parasuraman and Riley (1997) argue that the decision not to use automation has played a role in a number of accidents and resulted in monitoring problems.

Over a longer period of time, decisions about whether or not to use automation (and perhaps about whether to trust automation) have the potential to seriously impact on both skills and on task performance. A graphic by Llinas, Bisantz, et al. (1998) describes the theoretical consequences of what they call overtrust and undertrust.[7]

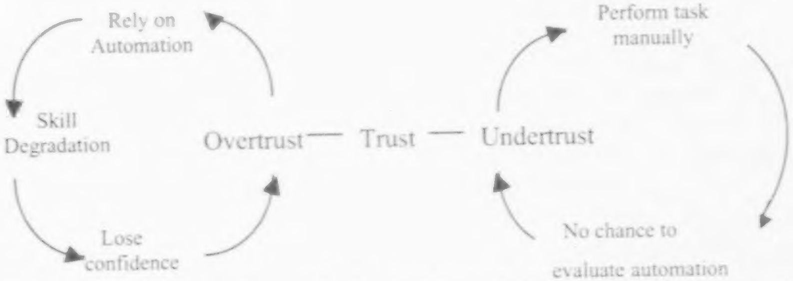


Figure 15: Vicious cycle of trust (adapted from Llinas, Bisantz, et al., 1998)

If an operator overtrusts a specific automated system, he is likely to rely on it more than is necessary given the system's reliability, thereby losing the skill to perform the task manually. This has been referred to as the *out-of-the-loop problem*, wherein use of automation essentially puts the operator out of touch with the performance of the system (Endsley & Kiris, 1995). This loss of skill can result in loss of confidence that may create a bias to continue using the automation. Conversely, if an operator undertrusts an automated system, he is more likely to perform the task manually as opposed to taking advantage of the benefit of the automation. As a result, there is no chance to re-evaluate the automation since it is not used. Undertrust is more resistant to change than trust, in that operators are prevented from gathering further evidence disconfirming that may disconfirm their expectations about its behaviour.[8]

From a broader perspective, then, both overuse and underuse of automation perpetuate the failure to calibrate one's expectations, beliefs and behaviours toward the true properties of an automated system. Calibration can only occur in response to new evidence, evidence that may change one's prevailing strategy and dictate a different response to the automation. Without relying on the automation, no new evidence that speaks to the reliability of the automation is possible. By relying solely on the automation, skills are likely to degrade, making automation less likely to be used properly, and more likely to show poor performance. In each of these cases, calibration of one's behaviour and expectations to the true performance of the automation is also precluded.

8.3 Trust and Monitoring

In the previous section, we established that use of an automated system can be predicated, to a certain extent, on an operator's level of trust in the system. This section focuses on the relationship between an operator's trust in automation and the monitoring of automation.

---

[7] Again, we disagree with these terms, but the graphic contains important and relevant information.

[8] It is important to note that although the relationship between trust and reliance on automation is often understood in terms of trust leading to more reliance, the opposite can also be true. The use of automation may itself be a predictor of the development of trust, in accordance with self-perception theory (Bem, 1967).

In the interpersonal trust domain, theorists have argued that when trust between partners in a relationship is low, they will be more likely to observe the activities and performance of each other (McAllister, 1995). This is referred to as *defensive monitoring*. In the context of trust in automation, defensive monitoring reduces the probability of potentially negative consequences when the automation does not conform to expectations and predictions.

The term "complacency" is used frequently in the literature to describe the failure to properly monitor automation.[9] Kelly et al., (2001) suggest that operators are complacent when they fail to monitor an automated system sufficiently so as to detect faults. That is, they are insufficiently vigilant. Similarly, Parasuraman and Riley (1997) state that complacency involves an overreliance or inappropriately low level of suspicion about an automated system that often results in a failure to appropriately monitor the system. Complacency has been implicated in a number of accidents involving aircraft and other high-technology systems (Singh et al., 1993). For example, Mosier et al. (1994; cited in Parasuraman & Riley, 1997) found that 77% of ASRS (aviation safety reporting system) incidents in which overreliance on automation was suspected involved a probable failure in monitoring.

On the other hand, the assumption within the literature is that complacency is *caused* by trust. This has not yet been widely shown. Although there has been much theorizing about complacency, and to a lesser extent, about skeptical monitoring, there is little empirical data that provides compelling evidence of these issues (Moray & Inagaki, 1999). It is, unfortunately, impossible to judge the extent to which these purported effects can be directly attributable to trust, as trust is typically not measured. As noted earlier, the relationship between complacency and trust is also problematic, in the sense that a lack of monitoring can clearly be due to high trust, but may also occur for reasons other than trust.

Nonetheless, it is clear that although monitoring is not necessarily related to trust in automation in all cases, there is a relationship between trust and some forms of monitoring behaviour. Muir (1994), for example, theorized that monitoring behaviour is a U-shaped function. Automation that is highly trusted will be monitored infrequently because uncertainty is low. Similarly, automation that is highly distrusted will require less monitoring, because operators will likely override distrusted automation and perform the task manually. Middle values of trust will be associated with high uncertainty and result in close monitoring of that automation. Muir predicted that if operators must continue to use distrusted automation, then they will continue to monitor its behaviour closely. It is also likely that operators will monitor all new or unfamiliar automation until they have gathered enough evidence to reduce their uncertainty enough to adopt an expectation of trust or distrust toward it.

Other research does provide some direct evidence that there is a relationship between trust and monitoring. Work by Muir and Moray (1996), for example, found that "the less operators trusted an automatic pump, the more intensely they monitored it and the more they trusted an automatic pump, the less intensely they monitored it". This finding is an important one, in the sense that trust in the automation as a psychological state was actually measured. This work suggests that trust can be related to monitoring. When trust is high, defensive monitoring can be low.

As noted earlier, however, the desired relationship between trust and monitoring can only fully be understood if the optimal rate at which automation should be monitored can be objectively defined. Moray and Inagaki (1999), for example, have defined three classes of monitoring behaviour:

1. Complacent – operators monitor less often (i.e., undersample) than is indicated by a model of the optimal observer

---

[9] Complacency is conceptualized interchangeably as overuse of automation, failure to monitor automation, and lack of vigilance directly due to trust in automation. This discussion of complacency relates only to the monitoring of automation.

2. Skeptical – operators oversample, thus wasting time on unnecessary observations which could be used for other aspects of plant management
3. Eutactic – operators sample at optimal rate

As Moray and Inagaki (1999) argue, unless one is able to determine the optimal sampling frequency or rate at which to monitor automation (i.e. eutactic monitoring), there can be no evidence that users may potentially overtrust or undertrust a system. Put another way, even if trust as a psychological state could be established as a potential factor in diminished monitoring behaviour, the issue of the optimal frequency of monitoring still needs to be resolved. Further, Moray and Inagaki (1999) point out that even if monitoring rate were eutactic, not all abnormal signals would be detected, given the attentional limitations of operators. As a result, there is a "non-zero probability that a fault will appear on a channel which is not at that moment being monitored." They argue, however, that this does not mean that operators are not monitoring at an optimal rate. In order to argue that monitoring is truly complacent because of trust, it would be necessary to show the presence of trust, articulate what the optimal rate of monitoring is, and be able to calculate the difference between these two factors. In attempting to understand the relation between defensive monitoring and trust in automation, then, it will be critical to make a distinction between an optimal sampling rate of monitoring (given that most forms of automation require some degree of monitoring) and monitoring which is purely defensive.

In contrast to complacency, *skeptical monitoring* is often equated with an operator's lack of trust in an automated system, resulting in an increased and unnecessary rate of monitoring. Operators who do not trust a reliable system may waste resources by monitoring the system more than is necessary. Skeptical monitoring has not received as much attention in the literature as complacency. This is likely because the consequences of skeptical monitoring, such as wasting time, appear to be less detrimental than those related to complacency.

Clearly, although there appears to be a relationship between trust and monitoring, both very high and very low levels of monitoring have the potential to be problematic. Too much monitoring has the potential to lower the efficiency of performance, and to distract operators from attending to other issues. Too little monitoring, on the other hand, could create a situation in which the operator loses touch with how the system is performing. If unexpected events occur, this could also present a serious problem.

8.4 Trust and System Performance

There are several theoretical accounts related to the impact of trust in automation on system performance. In 1994, Muir proposed that appropriate trust and appropriate distrust results in optimal system performance (see Table 8).

Table 8: How the operator's trust in and use of automation interact with the quality of the automation to influence system performance.

Operator's trust and allocation of function	Quality of Automation	
	'Good'	'Poor'
Trusts and uses the automation	*Appropriate trust* optimize system performance	*False trust* risk automated disaster
Distrusts and rejects the automation	*False distrust*	*Appropriate distrust*

	Lose benefits of automation, increase operator's workload, risk human error	optimize system performance

On the other hand, both false trust and false distrust are argued to result in non-optimal system performance.

Muir and Moray (1996) used a simulated milk pasteurization plant to, first, investigate the nature and dynamics of human trust in machines and second, to explore the relationship between the properties of the automation, trust and human intervention. In the first experiment, they found a positive correlation between performance measures and operator subjective ratings of trust. In the second experiment, however, they found that performance scores did not correlate with operator trust, indicating that operators did not base their trust on system performance.

Lee and Moray (1992) investigated the effects of varying frequencies and magnitudes of faults in a simulated orange juice pasteurization plant on operator's subjective level of trust as well as objective system performance. They found that both trust and productivity decreased with the introduction of faults but then gradually recovered toward pre-fault levels. Trust, however, did not recover as quickly as productivity, thereby demonstrating that recovery of trust lags system performance. In addition, they found that unlike the effect of faults on trust, the magnitude of the drop in performance did not correspond to the magnitude of the fault.

Finally, Lewandowsky et al. (2000) collected both subjective ratings of trust and objective measures of plant efficiency and reliance on automation. Participants were led to believe that they were working with either a human or a computer to control functioning of a dynamic simulated industrial plant. Automation faults were introduced during both auxiliary (i.e. automation in control) and manual (i.e. humans in control) conditions. When these faults occurred in auxiliary mode, trust in the other human operator and plant efficiency decreased. However, it was found that neither operator trust nor system performance were affected when faults occurred under manual control, and was true regardless of whether participants believed they were working with another human or automation. This suggests that the relationship between trust in automation and overall system efficiency varies depending on the underlying cause of faults.

Once again, although empirical studies have shown a generic relationship between trust in automation and overall system performance, this issue has yet to be adequately explored. This is an important area for future work.

8.5 Improving Trust in Automation

The literature provides some direction about ways to minimize the negative consequences of trust in automation. To do this, one must attempt to minimize the uncertainty associated with an automated system so that the operator perceives it as more predictable. This may be done in several ways, as noted below:

First, it has been shown that advance knowledge of system faults can reduce the uncertainty of the system thereby increasing levels of operator trust. For example, Riley (1994; 1996) established that one might be able to mitigate adverse effects of system faults on trust if operators have prior knowledge of magnitude and extent of failure.

Second, by making automation behaviours and system states more salient, the operator is likely to have more certainty about the system and therefore more trust in it. Further, more salient system behaviours can minimize the attentional demands of monitoring, thereby allowing the operator to attend to other aspects of the system. The salience of automation can be increased by way of direct perception displays and emergent features.

Third, Parasuraman et al. (1994) claim that providing adaptive automation with interspersed automated and manual tasks may help prevent to complacency or overreliance on automation. Intentionally introducing manual tasks at regular intervals, for example, may help to maintain an operator's vigilance. Varying the reliability of an automated system may also promote maintained vigilance on the part of operators (Parasuraman et al., 1993).

Finally, in order to prevent operators turning off alarms or automation due to undertrust, designers must take into account not only decision thresholds of the automation but also a priori probabilities of the condition to be detected (Parasuraman et al., 1994). That is, the actual probability of an event, compared to the number of alarms signaling this event must be determined. This will promote a more reliable basis for trust in the automation.

8.6 Overview and Research Implications

Although it seems clear that trust in automation can affect the extent to which one relies on and monitors the automation, these assumptions have been to be fully tested. Moreover, the extent to which an operator's trust in automation impacts on the actual performance of the operator/automation system has also yet to be clearly shown. The impact of trust in automation, in terms of both process and observable performance, need to be explored in more detail. Before doing this, however, it will be important to resolve some of the conceptual ambiguity in the literature around issues of trust, distrust, overtrust and undertrust. We would argue that the best way to do this may be to focus on the relevant behaviour (e.g. low rates of monitoring), rather to make assumptions about the underlying psychological state associated with these behaviours. Only then can the effects of trust in automation be discriminated from other related but distinct constructs. The research implications of this section are:

1. Progress in understanding the consequences of trust in automation seems to have been negatively affected by conceptual confusion. Making future progress will require consistent and objective definitions of the dimensions related to trust in automation, and to the use of automation.
2. Empirical studies that do exist have used solely supervisory control or process control automation. As such, the generalizability of the results to other types of automation is questionable. Given that the focus of this literature review is on trust in automation within the military, it is imperative to, first, establish the extent to which existing empirical research is applicable to the types of automation encountered in this domain. Second, research on the consequences of trust in automation must be extended within the military context.
3. In order to understand trust in automation, it will be important to work toward understanding the optimal frequency of use and sampling rate of automation (e.g. monitoring) to the highest extent possible (Moray & Inagaki, 1999).

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 9 - Preliminary Model of Trust in Automation

Based on the results of this review and our understanding of the trust literature, we created a preliminary model of trust in automation. This model describes both the factors that impact on the development of trust in automation, as well as the process by which trust in automation is formed.

The conceptualisation of trust that guides this work is best expressed in the following definition of trust, originally created in the process of developing a model of trust development in small military teams (Adams & Webb, 2003):

> *Trust is a psychological state involving positive confident expectations and willingness to act on the basis of these expectations. Issues of trust arise in contexts that involve risk, vulnerability, uncertainty and interdependence. Trust expectations are created primarily by the interaction of the perceived qualities of the trustee and the contextual factors in play when trust decisions are made.*

Although created for a different context, the definition is still applicable to the domain of trust in automation, even though the referent (i.e. human vs. automation) is obviously different. This definition fully expresses the importance of contextual factors in judgements of trust, as well as identifying how trust expectations come into being. The last part of the definition argues that trust expectations are mainly a product of the interaction between perceived qualities of the trustee (in this case, the automation) and the contextual factors in play when decisions to trust are made. The full preliminary model is presented in Figure 16.

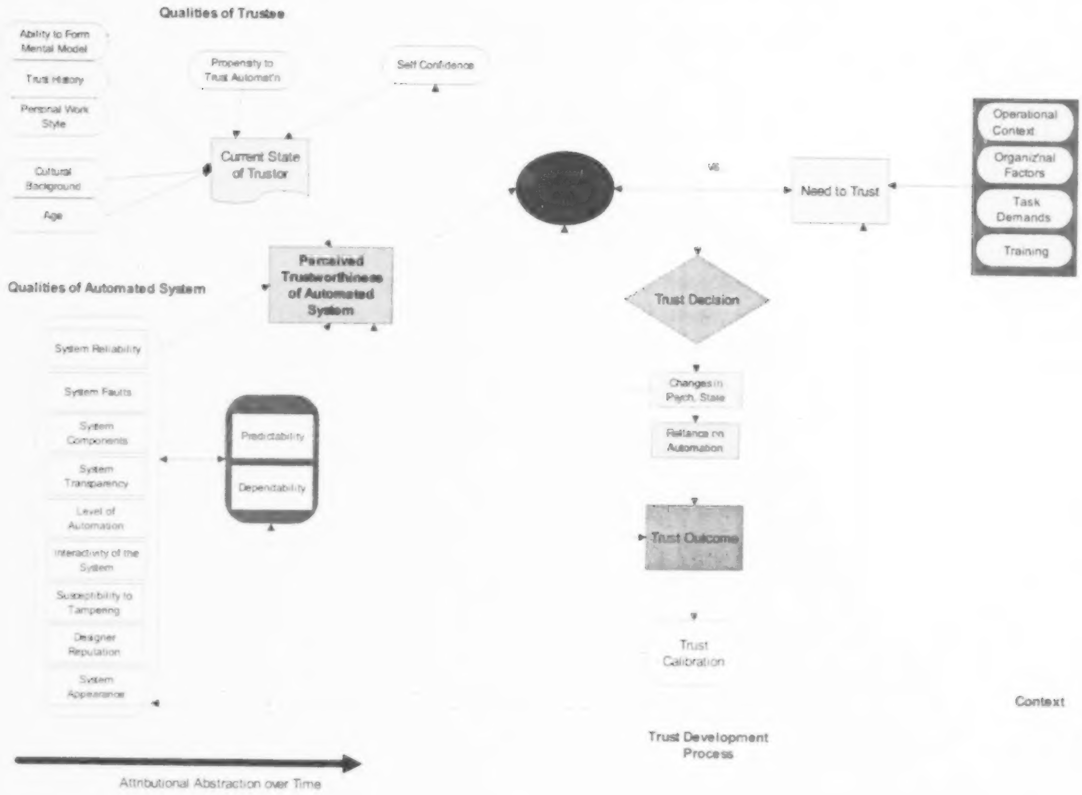

Qualities of Trustee
Ability to Form Mental Model
Trust History
Personal Work Style
Cultural Background
Age
Propensity to Trust Automat'n
Self Confidence
Current State of Trustor
Perceived Trustworthiness of Automated System
Qualities of Automated System
System Reliability
System Faults
System Components
System Transparency
Level of Automation
Interactivity of the System
Susceptibility to Tampering
Designer Reputation
System Appearance
Predictability
Dependability
vs
Need to Trust
Operational Context
Organiz'nal Factors
Task Demands
Training
Trust Decision
Changes in Psych. State
Reliance on Automation
Trust Outcome
Trust Calibration
Context
Trust Development Process
Attributional Abstraction over Time

Figure 16: Preliminary Proposed Model of Trust in Automation in Military Context

At the far left of the model, properties of the automated system are hypothesized to play the primary role in the development of trust in automation. The key variable in this model is the perceived trustworthiness of an automated system, and this is influenced by several sets of factors.

The first set at the upper left addresses the current state of person deciding whether or not to trust automation or the trustor. The current state of a trustor is influenced by a general propensity to trust automation. Age, cultural background, and ability to form a mental model of automation are also likely to influence trust in automation decisions. Self-confidence is also noted prominently in the literature as likely to influence trust in and use of automation. All of these factors combine to determine the current state of a prospective trustor interacting with automation.

Several qualities of an automated system are argued to influence trust in automation. These qualities are divided into two sets, actual properties and what may be called more emergent properties (predictability and dependability). The distinction between the various qualities on the far left and predictability and dependability indicates the fact that predictability and dependability develop over time as the result of attributions. As such, these factors are less properties of the system, but speak more to the interpretation of the system over time. Research clearly suggests that the major influence on the development of trust is the actual reliability of the system. This refers to the extent to which the system performs the task(s) that it was designed to do consistently over time. The existence of discrete events called faults is also closely inversely but not totally overlapping with the reliability of

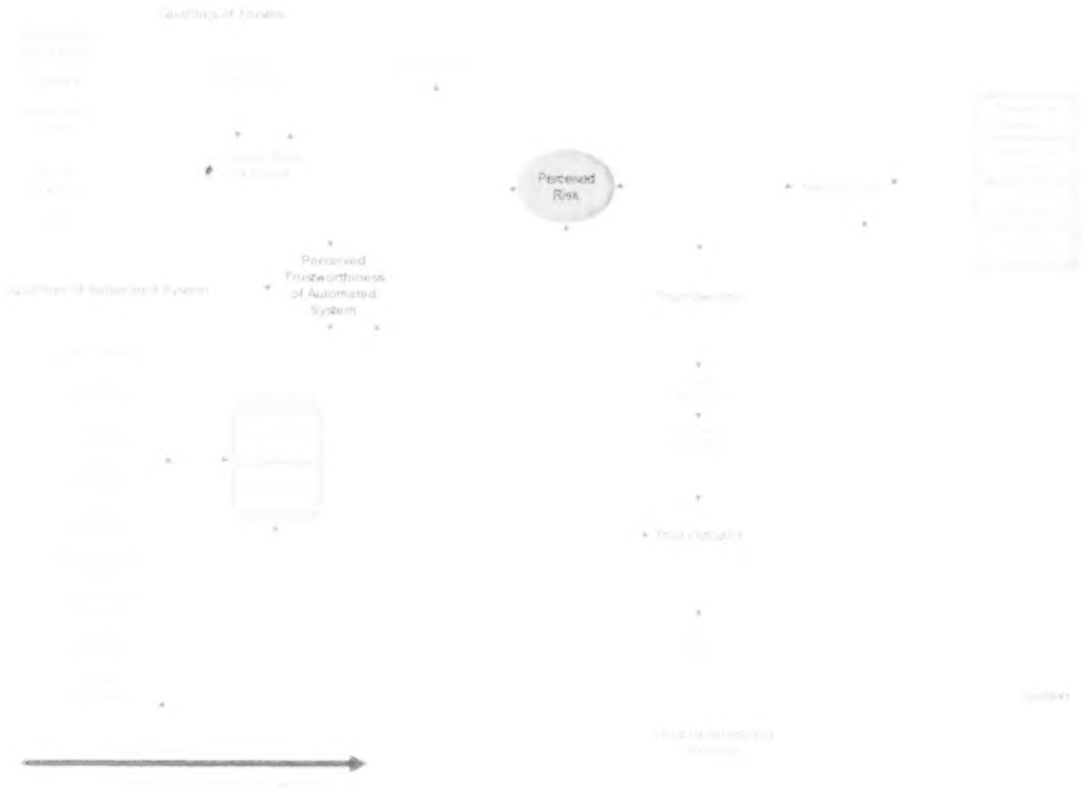

Qualities of Trustor
Perceived Risk
Perceived Trustworthiness of Automated System
Qualities of Automated System
Trust Decision
Trust Outcome
Context

Figure 16: Preliminary Proposed Model of Trust in Automation in Military Context

At the far left of the model, properties of the automated system are hypothesized to play the primary role in the development of trust in automation. The key variable in this model is the perceived trustworthiness of an automated system, and this is influenced by several sets of factors.

The first set at the upper left addresses the current state of person deciding whether or not to trust automation or the trustor. The current state of a trustor is influenced by a general propensity to trust automation. Age, cultural background, and ability to form a mental model of automation are also likely to influence trust in automation decisions. Self-confidence is also noted prominently in the literature as likely to influence trust in and use of automation. All of these factors combine to determine the current state of a prospective trustor interacting with automation.

Several qualities of an automated system are argued to influence trust in automation. These qualities are divided into two sets: actual properties and what may be called more emergent properties (predictability and dependability). The distinction between the various qualities on the far left and predictability and dependability indicates the fact that predictability and dependability develop over time as the result of attributions. As such these factors are less properties of the system, but speak more to the interpretation of the system over time. Research clearly suggests that the major influence on the development of trust is the actual reliability of the system. This refers to the extent to which the system performs the task(s) that it was designed to do consistently over time. The existence of discrete events called faults is also closely inversely but not totally overlapping with the reliability of an automated system. System components will also influence the development of trust, as trust in the display, control and the algorithms also guide trust expectations. The level of automation is also proposed to influence the development of trust. Automation that takes over more functions, on average, will be more difficult to predict. This makes it more difficult to trust. Automation also varies in terms of its transparency, or the extent to which its functioning, status, and output is both available and easily interpretable to the average user. Automation that is more transparent is likely to be seen as more trustworthy. The extent to which automation is seen as susceptible to tampering is a critical consideration within a military context. Automation that is more resistant to tampering will be more likely to be trusted.

To this point, the factors proposed to influence trust in automation correspond to what the interpersonal trust literature has called person-based factors. Within the automation context, these factors may be called target-based. We would also argue that within the trust in automation domain, category-based factors, factors predicated on the information associated with previously developed categories will also impact on the development of trust in automation. Even in the absence of direct evidence with an automated system, human operators may be able to quickly understand a new automated system, by using information relevant to the designer of the system. In this sense, the reputation of the designer of an automated system may have immediate impact on whether the automation is trusted. This reputation, of course, is expressed in the power of brand names, and in the fact that brand categories do, over time, come to carry implications for the trustworthiness of associated products. The brand name "Edsel", for example, is clearly associated with questionable trustworthiness.

Moreover, over time, attributions of predictability and dependability will also influence trust in automation. To the extent that an automated device fulfils one's expectations, it may be seen as predictable within this limited context. Over time, these attributions of predictability may come to extend to a broader level of attribution. At this point, the automation is no longer predictable within a specific domain, but its predictable behaviour comes to be understood as the product of a broader set of expectations about its underlying nature, or its dependability.

As indicated on the far right of Figure 11, the development of trust in automation will also be influenced by several contextual factors. First, trust in argued to become an issue only in the presence of the situational antecedents of risk, vulnerability, uncertainty and interdependence. In order for trust in automation to become an issue, there must be a need to work interdependently, to be unable to complete one's assigned task without the use of automation. The need to use automation makes one vulnerable to undesirable outcomes and to uncertainty, in the sense that the efforts to work together with the automation may or may not be successful. Within such situations, issues of trust come into play. Within a military context, however, trust in automation is also likely to be affected by the operational context. In the course of actual operations, for example, trust in automation is likely to be tested in a way that is not the case in garrison. Levels of risk and uncertainty in actual operations, for example, are likely to increase the need to trust automation. Similarly, the task at hand will influence the trust development process. Some forms of automation are particularly adept at helping with certain kinds of tasks (e.g. simple process control tasks), but it is more challenging to create a decision aid with the same absolute reliability of a simpler automated device. Organizational context will also greatly influence trust in automation. To the extent that an organization mandates automation use, and creates regulations governing its use, these regulations provide both explicit and implicit messages about the trustworthiness of the automation. A highly developed set of safety procedures, for example, may provide very clear information about the inherent trustworthiness of the automation. Training factors such as the quality of training and the amount of experience with the automation will also influence the development of trust.

As such, we argue that although the antecedents of trust in automation are different from those in the interpersonal domain, the actual process by which trust develops is very similar to trust development in the interpersonal domain. At the first level is perceived risk vs. the need to trust. An individual faced with a trust dilemma is faced with perceived risk on one hand and the need to trust on the other. Perceived risk can be seen as a relatively more internal process which is specific to the individual. High levels of perceived risk will obviously make a trust decision a more pressing need. The need to trust, on the other hand, is conceptualized as more of an external factor. The need to trust is cast as the product of several external influences or contextual factors including the operational, team, and organizational context. Within a given organization, for example, the need to trust is dictated by the demand to work cooperatively. Given a work situation in which the pressure to complete a given task demands interdependency, the need to trust is much higher. At the same time, however, perceived risk is also prominent. In a sense, then, decisions about whether to trust automation represent a "balancing act", in which the individual must gauge both their perceived level of risk in the situation and the pressure to enact trust from external forces.

Tension between the need to trust and perceived risks of trusting leads to a trust decision. In the trust literature, this decision is often called risk taking in the relationship (e.g. Mayer et al., 1995). As noted earlier, this decision may be manifested in either changes in trust at a psychological level, for example, by altering one's attitudes and beliefs about the automation, or as an observable choice behaviour which indicates trust. It is important to point out, consistent with the assertion that trust as a psychological state should take precedence over trust as choice behaviour, we see trust in automation first and foremost, as a psychological state first. This psychological state may or may not have implications for trust behaviour relevant to the automation.

Judgements about the trustworthiness of automation may be more at an implicit than explicit level. One may have the feeling that an automated system is trustworthy, but does not necessarily make a discrete and conscious trust decision in the course of interacting with the automation, as the situation may not demand such a decision. This model does not distinguish between these different forms of trust decisions, but argues that a trust decision may occur as either a discrete decision, or as a general impression that may be expressed in both implicit and explicit forms of trust decisions. The trust in automation model also depicts a trust decision as leading to a trust outcome. Trust outcomes tend to vary in their discreteness, and the intention is not to imply that every trust decision will necessarily lead to an immediate, discrete trust outcome.

Trust development is also depicted as having a feedback loop component, in the sense that the outcome of a trust decision influences subsequent trust expectations and possibly the likelihood of trusting behaviour occurring in the future. These outcomes, obviously, can be either positive (e.g., trust is rewarded) or negative (e.g., trust is violated). A trust calibration process can then occur. At this stage, the individual compares the result of the trust decision with the expected result, and makes a decision about whether or not to adjust the various factors which impact on their trust. Trust calibration could occur in many different ways, as indicated by the dotted lines in Figure 15. Potential trust calibration processes include:

- Perceived trustworthiness – When the outcome of trusting automation is positive (i.e. trust is rewarded), for example, perceived trustworthiness is likely to rise accordingly.
- Trust-relevant qualities– A trust outcome may result in an individual revisiting the source of trust expectations, the perceived qualities of the automation. This may include, for example, reassessing the perceived competence of the automation. Revising one's views of these qualities may result in better-calibrated trust expectations.

- Perceived risk – When trust is rewarded, the perceived risk inherent in future trust judgements is likely to diminish.
- Need to trust – As the result of trust outcomes, the external pressure to enact trust may be recalibrated. This may occur, for example, as the result of more accurately understanding the trust requirements within a given context.
- Current state of trustor – A trust outcome with serious consequences or which greatly violates one's expectations may lead to revisions in one's propensity to trust as a personal characteristic. This may also impact on one's self-confidence.

Of course, not all of these trust calibration processes will necessarily occur each time, and the outcome of the trust decision will influence which elements need to be recalibrated. In the case of a negative trust outcome, for example, one can attribute the failure to the automation, or to misreading the requirements within the context that the trust decision occurred.

The relationship between these sets of qualities can be seen as properties on one side (relevant to automation, the trustor's current state etc.) combined with the contextual factors. We have chosen to depict this relationship using an interactive function, on the basis that it most closely parallels the psychological processes in which people engage when making decisions to trust automation. Believing that automation is competent to perform a given task, for example, will only affect judgements of trustworthiness if the trust decision at hand implicates the known task. If the trust decision at hand does not require this specific form of competence, one may confer trust very differently. The depiction of an interactive function, of course, between the person and the situation follows on a common distinction prominent in social psychology (Alcock, Carment & Sadava, 2001).

Theoretical accounts of trust development have noted the importance of the relationship between the person and the situation (Mayer et al., 1995), but this perspective has yet to be fully incorporated into existing models of trust. This model attempts to do so, by specifying that trust expectations are determined by the interaction of a trustor, the automation's qualities and characteristics, and the features of a situation that are in place at that time.

Lastly, this preliminary model also addresses both the factors that influence trust in automation, as well as indicating the process by which trust develops. Moreover, this model applies equally well to different forms of automation, although different weightings of the importance of each factor are likely to differ. In the case of automation designed for process control, for example, a person's ability to form a mental model, or a system's transparency may be much less applicable than is the case with decision aids. This model, although preliminary, presents a complete account of both the factors that influence the development of trust in automation, and the process by which these factors exert influence. Further development and refinement of this model should be considered as part of a larger program of research. Measures of trust in automation and the features of a proposed research program are considered in the next two chapters.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 10 – MEASURES OF TRUST IN AUTOMATION

In this section, we describe approaches that may be useful in exploring trust in automation. Then, we present specific methods and measures for the study of trust in automation.

10.1 Research Approaches

Several different research approaches have been employed to investigate either trust in automation or very closely related issues (e.g. human-automation coordination). Several relevant approaches are discussed in this section.

10.1.1 Field Study of Real or Simulated Systems

Research has been conducted in natural settings such as in aircraft (e.g., Sarter & Wood, 1997) and in industrial plants (e.g., Zuboff, 1988). For example, Sarter and Wood (1997) employed the Airbus A-320 as a natural laboratory to study human-automation coordination. Field studies increase the realism of the experimental setting but allow less control over potential extraneous variables.

10.1.2 Simulators

Technological evolution now permits the use of simulators that provide high fidelity physical and dynamic replicas of real systems (Moray & Inagaki, 1999) and future research should take advantage of the high degree of realism and control that simulators offer. This technology is already utilized for training and research in aviation, navigation, and nuclear power plant control systems.

10.1.3 Microworlds

Microworlds are small-scale replications of process control systems designed to capture some of the complexity of the actual work domain, but intended to provide greater experimental control (Lee & Moray, 1994). They are designed to "… include closed loop coupling of humans and automated control in real time but are not general simulation of any real system" (Moray & Inagaki, 1999). Overall, this research approach represents a trade-off between highly controlled laboratory studies and field studies where having full experimental control can be more difficult. Microworlds have been used frequently by researchers (e.g., Muir & Moray, 1996; Moray, Inagaki, & Itoh, 2000) and, to date, represent the most common way that trust in automation has been studied.

10.1.4 Interviews

Interviews have also been used in order to understand trust in automation. In work by Simpson and Brander (1995; in Aldern, 1995), for example, semi-structured interviews were used in conjunction with other measures in a study exploring the Data Fusion Technology Demonstrator System, a prototype Command and Control system used in a naval context. Interviews are likely to be useful in understanding trust in automation, particularly at the early stages of research.

10.1.5 Experimental Research

Other than the use of microworlds, relatively little experimental research related to trust in automation has been conducted in the laboratory. Laboratory tasks provide a highly controlled experimental environment. Riley (1994) used simple computer games and a gambling task to provide empirical validation of his model of automation use. As research in the interpersonal trust domain has shown, it is possible to manipulate trust experimentally, and to look at the impact of either high or low trust on variables of interest (e.g. Dirks, 1999). Moreover, as we have argued, trust in automation is likely to be affected by contextual variables, such as high or low levels of risk and it is possible to manipulate these variables and explore the impact on trust as a psychological state and as choice behaviour. The use of experimental research, then, holds considerable promise for a better understanding of trust in automation.

10.1.6 Criteria, Measures and Methods

Based on a framework used by Matthews, Webb and Bryant (1999), we have identified measures and methods in the literature applicable to trust in automation with respect to three elements. First, a *criterion* identifies a broad dimension of interest. Specific *measures* and *methods* are then used to specify and operationalize the criterion. In many cases, there are multiple measures and methods applicable to a given criterion that provide differing levels of diagnostic power, or that are applicable in different contexts. Within any system of evaluation, a criterion may be fairly constant, but the measures and the methods used to capture the dimension of interest may vary depending on the need for precision and the availability of resources. Finally, a *standard* by which to interpret the data must be chosen. Standards cannot be chosen at this stage as they will vary depending on the experimental context chosen and on the demands of the experimental situation.

In the remainder of this section, we suggest criteria, methods and measures for different aspects of trust in automation reviewed in earlier sections. This is done with respect to the measurement of trust in automation as both a psychological state and as a choice behaviour which is associated with the psychological state.

10.2 Trust in Automation as a Psychological State

Trust as a psychological state has been measured in many different ways. In this section, we describe a few examples of informal measures of trust in automation as a psychological state that have been used in existing research, before turning to more complex efforts to measure trust in automation.

Work by Muir (1994), for example, explored trust in three aspects of a process control pump. This work addressed trust in general, as well as in respect of the following constituent parts:

1) trust in the pump to respond accurately,
2) trust in the pump's display,
3) overall trust in the pump.

In addition, these aspects were also rated on six factors underlying trust in automation: competence, predictability, dependability, responsibility, reliability over time, and faith in future ability.

Other work by Lee and Moray (1994) explored the psychological state of trust in various components of the automated system (e.g. feedstock pump, steam pump, steam heater), as well as confidence in the user's ability to control the system, as indicated in Table 9 below:

Table 9: Lee and Moray (1994) Subjective Rating Scale Statements

Questions:
1) How high was your self confidence in controlling the feedstock pump?
2) How much did you trust the automatic controller of the feedstock pump?
1) How high was your self confidence in controlling the steam pump?
2) How much did you trust the automatic controller of the steam pump?
1) How high was your self confidence in controlling the steam heater?
2) How much did you trust the automatic controller of the steam heater?

Scale items either refer to trust in an entire system, or in parts of the system. Although these kinds of scales have helped to provide information about the nature of trust in automation, they are limited, since they have generally not been empirically derived or validated. Moreover, these types of scales have not been created within the context of a fully developed model of trust in automation, but are often intuitively derived based on researchers' theories about trust in automation. There have been some formal efforts to create scale measures of trust in automation, considered in the sections that follow.

10.2.1 Empirically Determined Trust in Automated Systems (Jian, Bisantz, & Drury, 2000).

Jian, Bisantz, and Drury (2000) have worked to develop an empirically derived scale to measure trust in automation. The purpose behind their research was fourfold. The first facet of their work consisted of conducting an empirical analysis of the different theoretical notions of trust and uncovering the suggested multidimensionality discussed in the literature. The second facet involved demonstrating the opposing nature of trust versus distrust. Their third objective was to explicitly evaluate how interpersonal trust differs from trust in automation. Finally, based on results obtained from these studies, the last goal was to develop a potentially more valid psychometric tool to measure people's trust in automation. Given these objectives, it is important at this point to review their methodology.

A three-phased empirical study was first conducted. The first phase of the experiment consisted of a word elicitation task study in which participants were asked to provide a written description of their understanding of the concepts of both trust and distrust with regard to trust in other people, trust in automation and trust in general. Participants were also asked to rate a set of 138 words associated with trust on the basis of whether they related positively or negatively to the three different types of trust stated above. The result of this first phase permitted the development of a final set of 112 trust-related words.

The second phase consisted of a questionnaire study in which the participants were asked to analyze each of the 112 words generated in phase one in terms of whether they relate to trust or distrust, again with regards to the three types of trust; interpersonal trust, trust in automation and trust in general. The goal of this phase was to determine whether the concepts of trust and distrust could be viewed as opposites, and to determine whether trust and distrust could be viewed as similar concepts across the three types of trust. The results indicated that trust and distrust were strongly negatively correlated. In addition, the pattern of ratings for the three types of trust were very similar.

The final phase consisted of a paired comparison study. The goal of this study was to gather data about the factor structure of the positive and negative trust-related words rated in phase 2. Participants were required to compare and rate 30 positively and negatively trust-related words. Using factor and cluster analysis techniques, a 12 factor structure was extracted to develop a multidimensional measure of trust between humans and automation (see Appendix A). The proposed measure of human-machine trust was comprised of 12 items derived from the examination of the clusters of words. Items are rated on a seven point Likert scale, ranging from 1= *not at all*, to 7 = *extremely intense*.

Although this work is compelling, the conclusions reached by the authors are somewhat in need of further exploration. For example, based on the fact that the patterns of ratings of the words for the three types of trust were similar, the authors argue that future work need not treat interpersonal trust, trust in automation and trust in general as fundamentally different concepts. A separate analysis, however, suggests that trust in these three contexts may not be the exactly the same. Regression analyses showed that although there were no significant differences between trust in general and human-machine trust, there were significant differences between general trust and human-human trust and human-human trust and human-machine trust. It is unclear why these differences would exist if people saw trust as very similar in all three contexts, and an explanation of this contradictory finding is not offered.

In addition, the authors' assertion that a high negative correlation between trust and distrust indicates that they are a single construct on a bipolar scale is in need of further empirical work. The fact that two constructs are negatively correlated does not necessarily mean that they load on the same underlying construct. The question of the dimensionality of trust might have been resolved more conclusively with confirmatory factor analysis, but this kind of analysis was presumably not performed. Most importantly, because this scale is relatively new, studies pertaining to the psychometric properties (e.g., reliability, validity) of the measure still need to be performed. Nonetheless, this work is an important contribution to the trust in automation literature, as it represents the most complete effort to measure trust in automation to date.

Overall, this measure represents the first attempt at empirically generating a scale to measure trust in automation. One important feature of this scale is that it was constructed with respect to trust in automated system generally, rather than trust in a specific system (Jian, Bisantz, & Drury, 2000). As such, this scale is likely to be indicative of a general propensity to trust automation. One important implication of this new measure is that it could serve as an anchor or potential conceptual baseline for the development of psychometric instruments to measure trust in specific automation.

10.2.2 Human-Computer Trust Instrument (Madsen & Gregor, 2000).

As noted in the model chapter, work by Madsen and Gregor (2000) provides both a model of trust in a decision aid, and working to establish a measure of trust in a decision aid. In this study, a group of subjects identified constructs that they believed would affect their level of trust in an intelligent decision aid. Following refinement and modification of the constructs and potential items, the instrument was reduced to five constructs (reliability, technical competence, understandability, faith, personal attachment). Items contained in the Human-Computer Trust (HCT) instrument are presented in Table 10 below.

Table 10: Scale items in final Human-Computer Trust instrument (Madsen & Gregor, 2000)

1. ***Perceived Reliability***
R1 – The system always provides the advice I require to make my decision.
R2 – The system performs reliably.
R3 – The system responds the same way under the same conditions at different times.
R4 – I can rely on the system to function properly.
R5 – The system analyzes problems consistently.
2. ***Perceived Technical Competence***
T1 – The system uses appropriate methods to reach decisions.
T2 – The system has sound knowledge about this type of problem built into it.
T3 – The advice the system produces is as good as that which a highly competent person could produce.
T4 – The system correctly uses the information I enter.
T5 – The system makes use of all the knowledge and information available to it to produce its solution to the problem.
3. ***Perceived Understandability***
U1 – I know what will happen the next time I use the system because I understand how it behaves.
U2 – I understand how the system will assist me with decisions I have to make.
U3 – Although I may not know exactly how the system works, I know how to use it to make decisions about the problem.
U4 – It is easy to follow what the system does.
U5 – I recognize what I should do to get the advice I need from the system the next time I use it.
4. ***Faith***
F1 – I believe advice from the system even when I don't know for certain that it is correct.
F2 – When I am uncertain about a decision I believe the system rather than myself.
F3 – If I am not sure about a decision, I have faith that the system will provide the best solution.
F4 – When the system gives unusual advice I am confident that the advice is correct.
F5 – Even if I have no reason to expect the system will be able to solve a difficult problem, I still feel certain that it will.
5. ***Personal Attachment***
P1 – I would feel a sense of loss if the system was unavailable and I could not longer use it.
P2 – I feel a sense of attachment to using the system.
P3 – I find the system suitable to my style of decision making.
P4 – I like using the system for decision making.
P5 – I have a personal preference for making decisions with the system.

A field study with users of operational taxi dispatch systems was used to test the instrument for construct validity and scale reliability. The average inter-rater reliability for the scale as a whole was very high ($\alpha = .94$). Principal components analyses were then conducted in order to understand relationships between the scale item ratings and the underlying construct, trust. The first analysis showed that all scale items were related to a single dimension, trust. Another analysis was then conducted limiting the final model to two dimensions, cognition-based trust and affective trust. Results were argued to show that both items loaded predictably on the predicted dimensions. Another analysis limiting the solution to a five factor model (the proposed factors) was then performed, and showed that most of the scale items were related to the proposed underlying factor.

This analytic strategy, unfortunately, does not allow for a fair assessment of the measurement instrument. We would argue that conducting multiple principal component analyses in separate analyses is not the most appropriate analytic strategy. The authors had clear hypotheses about the underlying relationships between the scale items, the proposed factors and the underlying variable, human-computer trust. As such, a superior approach would have been the use of confirmatory factor analysis, which would have allowed for a simultaneous test of the proposed relationships, rather than several separate analyses. To its credit, however, this measure is strongly conceptually grounded, an important contribution in and of itself.

10.2.3 Complacency Potential Rating Scale

Work by Singh et al., (1993) sought to create a scale to measure complacency potential, the extent to which an operator's attitude towards commonly encountered automated devices reflects a tendency to rely on automation. We have already argued that a conceptual distinction should be made between complacency and overtrust, or reliance on automation. Nonetheless, this scale does include an effort to measure trust as a psychological state in one of the subscales.

The Complacency Potential Rating Scale (CPRS) was developed in three stages. The first stage involved the generation of an initial pool of 100 items, intended to assess attitudes (favourable vs. unfavourable) towards automation. Four subject matter experts evaluated these statements. The reviewers selected a set of 20 items for the final version of the scale (see Appendix B) on the basis of face validity in assessing the suggested five factor structure of complacency: confidence, reliance, trust, safety, and general attitude regarding automation. These items were evaluated on a five-item Likert scale with anchors ranging from strongly agree to strongly disagree.

The second and third stage involved the assessment of the empirical properties (validity and reliability) of the scale. The scale was administered to 139 volunteer subjects. The favourable and unfavourable items were reverse scored, and composite scores on the measure were calculated from the addition of all item scores. Both reliability and factor analyses were conducted on the items in the scale. Inter-item correlations were computed for the scale. It was found that all of the items were highly correlated ($\underline{r} = 0.98$), indicating a high internal consistency among the items. An item discrimination analysis displayed an extremely low value ($\underline{r} < 0.22$), which supports the retention of all items. The test re-test reliability of the CPRS was also quite high ($\alpha = 0.87$). Finally, as expected, a five factor structure explaining about 53% of the variance was also found. Examination of these factors in their relative order of importance revealed that general attitudes concerning automation, confidence, reliance and safety accounted for about 19%, 11%, 9%, and 7% of the variance respectively. Most importantly for our purposes, trust in automation accorded for 8% of the variance in complacency potential.

Although this measure was found to be reliable and internally consistent, two major limitations are worth noting. First, additional work is required to determine its construct and discriminant validity (Singh et al., 1993). Secondly, looking at the scale items related to trust in automation, it seems that items such as "manually sorting through card catalogues is more reliable than computer-aided searches for finding items in a library" tap a very general tendency to trust automation, which would not necessarily be helpful for understanding trust in a specific automated system. In any case, this scale would have little use in the context of the current research.

10.2.4 SHAPE Automation Trust Index (SATI)

A report produced for the European Organisation for the Safety of Air Navigation (EUROCONTROL) discusses the development of a scale to measure trust in air traffic management (ATM) systems (Goillau, Kelly, Boardman, & Jeannot, 2001). After a review of the trust in automation literature (Kelly et al., 2001), the development of a measure of trust in air traffic management systems was undertaken. Human factors technique was used to develop a scale to be used primarily in real-time simulations of future ATM systems.

Based on the literature review, and on the conceptual model described earlier, preliminary measures of trust in ATM systems were designed. The scale is aimed at the "pragmatic measurement" of trust, and has been created through the use of several different modules depicted in Table 11.

Table 11: Components of SHAPE Automation Trust Index for trust in Air Traffic Management (Goillau et al., 2001).

Module	Detail
Overall amount of trust	In the simulated system / automation tool (on a % scale from 0% (no trust) to 100% (complete trust))
Variation over time	Trust level at the beginning and end of a time period
Decomposition of trust	Decomposition into dimensions of trust, with rating of the relative importance of each dimension - ratings of trust factors (e.g. reliability, accuracy, understanding of intention etc.)
Distribution of trust	Ratings of trust in the specific automated tool, the simulated system, self-confidence, local colleagues and trust in others.
General comments	Open ended responses on factors that influence trust and how trust in the system could be increased.

Any number of these modules could be administered throughout the course of the simulation or at the end. The most recent version of this scale is attached in Appendix C.

This measure was then refined in focus groups with air traffic controllers. More specifically, refinement of the measures occurred during two real time simulations with air traffic controllers from different cultures. In addition to the rating scales, usability evaluation trials, and construct validity feedback from air traffic controllers were also collected at this time.

A number of considerations emerged during the scale development and refinement processes. First, the scale was shown to be usable. Controllers knew what they were being asked, and the measures had some degree of face validity. From the controller perspective, trust was defined as "Reliable performance in terms of behaviour" and was regarded as binary factor (yes or no). Controllers argued that they either trusted the automation or did not, as indicated by the fact that 78% of controllers agreed with the statement "I either trust, or I do not trust, ATM automation". Confidence, on the other hand, was argued to be variable, and 78% of them agreed with the statement "But I have varying degrees of confidence in ATM automation". Unfortunately, there is no information that speaks to the definition of confidence, but it is argued to be distinct from trust, as it is more fine-grained.

As a whole, this work represents the kind of effort most in keeping with what would be necessary in measuring trust within the military domain. The measures are pragmatic and consistent with how trust is understood by the participants within the air traffic control domain. Although we do not necessarily agree with several conclusions made by the researchers (e.g. trust is binary construct), the researchers did actively work to create their scales to reflect how air traffic controllers understand trust. Lastly, the measurement instrument itself was created based on feedback from air traffic controllers, and is designed to be easy to use and understand. The use of graphics within the measurement instruments, for example, is unique among the other trust in automation efforts that we have reviewed.

Future measures of trust in automation need to address two different levels. First, the summary judgement of trust in automation that an individual reports is relevant. This is the level most commonly used in extant informal measures of trust in automation (e.g. "to what extent do you trust the system?"). Secondly, an index of the perceived trustworthiness based on the factors believed to influence trust in automation will also need to be created. In terms of the first level, trust in automation can be assessed as a whole (e.g. "to what extent do you trust the system?"), or in terms of cognitions, affect, and intentions toward the automated system, as indicated in Table 12 below.

Table 12: Measures and Methods for Assessing Trust in Automation as a Psychological State

Function	Criterion	Measures	Method
Trust	Positive expectancies	Trust-related expectancies	Participant ratings on predefined scales
	Positive feelings	Trust-related feelings	Participant ratings on predefined scales
	High willingness to trust	Willingness to trust	Participant ratings on predefined scales

As noted earlier, several different sets of factors are likely to influence trust in automation. In accordance with the literature, we would expect that properties of the automation are likely to be the strongest determinants of trust in automation. Several properties of automation with suggested measures and methods are listed in Table 13:

Table 13: Measures and Methods for Assessing Properties of the Automation

Function	Criterion	Measures	Method
System reliability	High system reliability	Efficiency, failure rates	System provided information Observer ratings Participant ratings on predefined scales
System faults	Minimal number of system faults	# of system faults variability, locus, magnitude, participant knowledge of system faults	System provided information Observer ratings Participant ratings on predefined scales
System components (display, control, algorithm)	High trust in system components	Trust in system component(s)	Participant ratings on predefined scales
System transparency	Transparency of output Transparency of feedback	Transparency of output and feedback	Observer ratings Participants' ratings on predefined scales
Level of automation	High level of automation	Sheridan/Verplank ratings	Comparison to Sheridan/Verplank framework
System interactivity	High interactivity with automation	Type and frequency of interactivity	Participant ratings Observer ratings SME ratings
Susceptibility to tampering	High protection against tampering	Ratings of safeguards designed to lessen tampering	SME ratings
Automation Designer	Good reputation of automation source	Ratings of source's reputation	Expert ratings Participant ratings on predefined scales (e.g. reputation of source)
Appearance of automation	Sturdy appearance	Ratings of appearance	Observer ratings Participant ratings
Predictability of automation	High predictability	Attributions of predictability	Participant ratings on predefined scales
Dependability of automation	High dependability	Attributions of dependability	Participant ratings on predefined scales

In addition, several properties of the trustor are also likely to impact on trust in automation. Among these, propensity to trust automation and self-confidence are likely to be the most influential. Ability to form a mental model of automation as well as cultural background are also likely to impact on trust in automation and should be considered in future measures of trust in automation, as indicated in Table 14:

Table 14: Properties of the Trustor

Function	Criterion	Measures	Method
Propensity to trust automation	High propensity to trust	Propensity to trust	Participant ratings Propensity to trust scale
Self-confidence	High self confidence	Self-confidence	Participant ratings on predefined scales
Ability to form mental model	Mental model of automation accurate and complete	Complexity and accuracy of mental model	Compare overlap with objective description of automation
Trust history	Positive trust histories	Past trust histories	Structured interviews
Personal work style	Match between personal work style and automation style	Ratings of preferred personal style SME reports on automation's style	Assess match between participant work style and automation configuration Participant ratings on predefined scales
Cultural background	Diverse cultural backgrounds	Individuals' cultural background	Participant ratings on predefined scales
Age	Experience with automation	Participant's age	Participant ratings on predefined scales

Properties of the environment will also influence trust in automation. The operational context in which decisions to trust automation occur will impact on trust. One might expect that under high stress operations, the need to trust automation (and hence the need to trust automation) will be much higher. Similarly, decisions about whether or not to trust automation also occur within a specific organizational context, in which rules and procedures may influence trust in automation. Features of the specific task at hand, and training related to automation will also influence trust, as indicated in Table 15:

Table 15: Properties of the Environment

Function	Criterion	Measures	Method
Operational Context	Extent to which operational context requires trust in automation	Garrison, field exercises vs. operations	Contextual analysis
Organizational factors	High organizational demands related to automation	Rules and regulations related to use of automation Organizational history and values	Contextual analysis
Task demands	High task complexity, workload and time pressure	Complexity and difficulty of tasks Level of operator workload Time demands	Observer ratings Participant ratings on predefined scale
Training	High quality training with automation	Kind of training (e.g. hands on experience) Length of training	Observer ratings Trainer ratings Participant ratings

10.3 Trust as Choice Behaviour

The existing literature suggests that trust as a psychological state may also be related to certain behaviours, as described below.

10.3.1 Defensive monitoring

Defensive monitoring is a trust-related behaviour that is undertaken to lessen the probability of potentially negative consequences if automation does not conform to expectations and predictions. It is argued by Muir and Moray (1996) that trust is negatively related to the monitoring of automation (Muir & Moray, 1996). If automation is highly reliable, defensive monitoring is less likely to be necessary. To understand the relation between defensive monitoring and automation, however, it will be critical to make a distinction between an optimal sampling rate of monitoring (as most forms of automation require some degree of monitoring) and monitoring which is purely defensive.

Measures

Self-report scales (measuring perceptions of how much monitoring is required) as well as behavioural measures of defensive monitoring could be employed to measure this behaviour. Defensive monitoring of automation could be measured, for example, in terms of the number of checks on the automated system, beyond the optimal frequency.

Methods

The proposed method for measuring defensive monitoring is the analysis of an audio-visual log by trained observers. This form of analysis provides observers with the time necessary to decode the information and to make decisions concerning the occurrence of defensive monitoring.

Table 16: Measures and Methods for Assessing Defensive Monitoring

Function	Criterion	Measures	Method
Defensive monitoring	Minimal defensive monitoring	Number of checks on automation above optimal rate	Analysis of Audio/visual log

10.3.2 Use of Automation

Although we have argued that the use of automation is not equivalent to trust in automation, if trust as a psychological state has been established, the use of the automation may indicate trust expectations in action. In some cases, then, the use of automation may be indicative of trust in the automation.

Measures

Automation use can be represented as the frequency of automation use and the length of each period of automation use. This can also be extended to the total proportion of time that the system is under automatic control vs. manual control.

Methods

The proposed method for measuring the use of automation is through a system log which collects data on automation vs. manual control. If this is not available, an audio/visual log could be analysed to determine the use of automation.

Table 17:

Function	Criterion	Measures	Method
Use of automation	Optimal use of automation	Frequency and length of automation use Time taken to act Types of information accessed during use of automation (e.g. manuals, help systems) Timing and sequence of decisions and actions	System output Analysis of Audio/visual log

10.4 Overview and Considerations for Measures of Trust in Automation

It is encouraging that trust in automation as a psychological state is considered important enough to measure, and that both formal and informal efforts have attempted to do so. The Jian et al. (2000) scale represents one notable effort that seems to measure a general propensity to trust, although even it has been criticized as being too emotive (Kelly et al., 2001). Moreover, the need for both valid and reliable scales is also commonly acknowledged to be important. To date, however, most scales have not been empirically derived, and few have been subject to any extended validation efforts. As such, we would argue that an adequate measure of trust in automation has not yet been developed. In developing such a measure, several issues will need to be considered.

First, future efforts to create measures of trust in automation should be grounded at the theoretical level in a conceptual model. In this review, we have proposed a preliminary model of trust in automation. This model should be refined further, and should serve as the conceptual base for future efforts to measure trust in automation within the military domain.

Secondly, any measure of trust in automation needs to be clearly defined from several different perspectives. We have argued that trust in automation is relevant at least two different levels, namely, propensity to trust automation in general, and trust in a specific form of automation or automated system. However, it may be impossible to create a single measurement instrument that can be meaningfully used for all levels of automation. The issues associated with trust in very simple automation, for example, are likely to be at least somewhat different from those involved in interactions with complex decision aids. In an ideal world, it would be possible to tailor a measure of trust in an automated system very specifically to the actual properties of the system. This, of course, limits the broader applicability of the measure, and creates problems for validating the measure. In the long run, in fact, it may be more appropriate to begin by developing a broad measure of automation and validate the scale within the domain in general, and then extend it via subscales to more specific forms of automation.

Thirdly, it is important to note that, in the military context, actually using the word "trust" in creating measures of trust in automation may not be particularly productive. For some people in the military, the concept of trust carries with it connotations that are not necessarily accurate because they focus too heavily on the affective and social aspects of trust. A tank crew commander who was part of an earlier focus group exploring trust in military teams argued that reliability (his preferred term) and trust were not entirely interchangeable:

> "*...I think it's the same terms, but in the military...trust is like something that's (SPEAKER MAKES A SCORNFUL FACE)...we count on reliability, can I rely on the guy...not trust him...because trusting is more like a feeling or touchy-feely kind of thing...but reliability - is that soldier reliable can I count on him...is he dependable*?" (Adams and Webb, 2003)

This aversion to thinking about trust (particularly in the context of automation) is likely to be common in the military and is consistent with the views of air traffic controllers who do not think in terms of "trusting" the system that they work with, but rather in terms of its operational reliability (Kelly et al., 2001). Avoiding or limiting use of the word "trust" in measuring trust in automation is not particularly problematic if we can find alternative terms that capture its essence and are agreed upon by users. The important issue here is in establishing whether users rate trust in automation low because they feel uncomfortable talking about trust (e.g. because of the commonly held belief that trust is only affective in nature) or, alternatively, whether trust is less of an issue in some types of relationships with automation. Although we would argue that the former is more likely to be true, it will be important to consider this issue in any future efforts to measure trust in automation.

Fourth, particularly in the latter stages of this research, study participants should be ideally representative of the target user population. More specifically, participants should be military personnel who are required to use automation in the context of performing their jobs. It will also be critical to control for various background factors that may impact on trust in automation, including propensity to trust automation, age, cultural background, ability to form relevant mental models, and most importantly, self-confidence. In this sense, the issue of self-confidence is likely to be a predictor of trust in automation, and it will be essential to have a measure of self-confidence that can be tracked over the course of time, in parallel with evolving judgments of trust. It will also be important to either control for differences in participants' military background and experience with automation.

Fifth, although self-reported measures of trust as a psychological state are likely to play a key role in future efforts to measure trust in automation, the many problems inherent in subjective measures are no less problematic within this domain of research. Errors and biases and social desirability concerns, for example, are still potential problems. Moreover, the extent to which participants are naturally in tune with their changing levels of trust in automation may be questionable. Although it seems natural to think about trust in other people, participants may be less adept at reflecting on their trust in automation. On the other hand, the available empirical work suggests that trust in automation can be measured by self-report and that the moment to moment changes in trust while interacting with automation are both meaningful, and to some extent, predictable.

Lastly, it is critical that the measure be matched with the military domain requirements as much as possible. There is a need to empirically develop adequate measures of trust in automation. The creation of questionnaires that can be used in high-stress, multiple demand environments will be important. In order to be useful to the military community, any future scale needs to be practical and easy to administer, with minimal labour.

CHAPTER 11 – PROPOSED RESEARCH PROGRAM

11.1 Overview and Research Considerations

The goal of this literature review was to identify critical issues in the area of trust in automation within military contexts. Factors that influence trust in automation, and the consequences of trust in automation have been considered. Models related to trust in automation have been reviewed and a preliminary model of trust in automation has been proposed. This review now considers research strategies and questions that could be used to study trust in automation within military contexts.

There is a need to develop valid and reliable psychometric instruments that capture the nature of trust in automation and its related antecedent factors. Measures of trust in automation may help to gauge impending problems associated with changes in trust in automation that may impact negatively on performance. The ideal is to be able to predict trust in automation from knowledge of the automation, the person and the context in which they interact. This, presumably, would also be reliably related to the actual use of automation. It is highly desirable that any measures be strongly grounded at a conceptual level, and that they stem from a broader understanding of how trust in automation should be represented within the military context. In an earlier chapter, we proposed a preliminary model that depicts how trust in automation develops. In so doing, we argued that the role of context in determining trust in automation has received some attention, but has yet to be seriously incorporated at a conceptual level into existing models of trust in automation. We would argue that understanding context is important, as seriously alters decisions to trust automation. Refinement of this model, followed by its validation in the context of a larger research effort should be an important focus of the future research program.

Trust in automation needs to be separated at a conceptual level from the external military requirement to use automation. Whether one trusts automation or not, it is often critical to use it. If there is an external requirement to use automation, a person's psychological state may or may not parallel their use of automation. In any case, if there is an external requirement to use automation, the issue of one's actual level of trust in the automation is, at least to some extent, a moot issue. In studying trust in automation, then, it is necessary to create contexts in which trust as a psychological state is primary. Only then can trust-related behaviour, such as the use of automation, be understood.

Before embarking on the experimental research program described in this review, we suggest that an initial broad, descriptive approach may be valuable. Such an approach may employ focus groups, surveys and questionnaires in order to explore soldiers' general perceptions about trust in automation. Within the military, however, it is not enough to consider the impact of conventional factors (e.g. system reliability etc.); issues such as who made the automation, and to what quality standards, are also likely to play a role.

A future program of research exploring trust in automation should be designed in such a way that, in some form, it helps to answer the question of how trust in automation can best be promoted. As such, it seems important to strike a balance between basic research exploring trust in automation and being able to give practical advice about how best to maximize trust in automation within the military context. At some point, for example, attention should also be paid to the issue of training, and to how training can be designed in order to promote the maximal possible level of trust in reliable automation. What kinds of knowledge promote trust in automation even before actual use? These kinds of questions have yet to be explored at an empirical level.

Perhaps more importantly, it is critical to go beyond the existing empirical work, both in terms of the methods used and in terms of the kinds of automation that have been explored. As noted earlier, the majority of empirical work has focused on trust within process control scenarios (or microworlds), in which participants are asked to control the functioning of an industrial plant through allocating either manual or automatic control. This context, although important for understanding function allocation, may or may not be indicative of the processes that people go through in making decisions about whether to rely on the recommendation of a decision aid. Within military contexts, there are many different forms of automation, involving both receiving advice and choosing whether to allocate tasks to oneself or to automation. As such, it seems important to extend beyond microworlds, and to broaden the kinds of automation explored through empirical work in order to be able to understand how trust in automation is likely to impact within a military context.

The next section poses several specific questions that might be asked in the context of the current research.

11.2 Proposed Research Approach

We believe that the key research focus for the issue of trust in automation should revolve around three main questions, as described below:

What is the nature of trust in automation?

We would argue that research into trust in automation should focus on trust as a psychological state, with trust as a choice behaviour treated as a secondary – but related – issue, thus avoiding the conceptual confusion in earlier research. However, it is important not to completely separate trust in automation from trust in the user of automation. A user must not only interact with automation, but often also needs to interpret the output and feedback of automation. Trust in both automation and in the user of automation will both need to be considered. At a broader level, it will also be important to understand how trust in automation is similar in structure and function to trust in other settings. For example, what are the key differences between trust in the interpersonal domain and trust in automation?

What factors develop and maintain trust in automation?

At a very practical level, it is important to understand how best to build and sustain trust in automation. The proposed model argues that many different sets of factors, the properties of automation, properties of the trustor, and properties of the context in which interactions with automation occur, will determine trust in automation. Some of these factors have been explored, and have received good empirical support. Others are still in need of validation.

As noted throughout this report, reliability of the automation is perhaps the critical factor influencing trust in automation. In this sense, understanding how to help people to optimally calibrate their levels of trust will be an important focus, in addition to considering the critical role of context and its impact on trust in automation.

The emphasis in the literature has been primarily on how trust in automation varies as the product of discrete system faults. Although the impact of faults is important, it will also be important to extend this research in order to understand the impact of events that violate the user's expectations about automation in a less discrete way. For example, it is possible for automation to fail to meet expectations in both negative and positive ways. The impact of automation unexpectedly performing better than normal, for example, is a topic worthy of further research.

The relationship between trust and monitoring of automation is also worthy of further examination. Although it seems reasonable that high trust is likely to be associated with relatively lower levels of monitoring, this is still in need of further investigation. The key issue here is being able to measure trust as a psychological state, and to quantify the degree of variation from the optimal level of monitoring (or indeed, any behaviour under focus). Monitoring can occur for reasons both related and unrelated to trust. How to conceptualize this distinction is a key challenge of future research.

Although this review has focused on the development of trust in automation, it will also be important to explore how trust is best maintained over time. The research suggests that trust violations (e.g. automation faults) pose a serious threat to trust in automation. It is important to understand what kinds of violations are most damaging to trust and how the negative impacts of these violations can be best mitigated in a human/automation environment.

How does trust in automation relate to performance?

The interest in trust in automation stems, in part, from the hypothesis that trust impacts positively on performance by improving effectiveness and efficiency. Interestingly enough, however, there is little empirical evidence that trust in automation actually improves performance directly. It might also be argued, however, that another way in which trust in automation is likely to have a positive impact is through psychological well-being. In theory, this should be reflected in measures of perceived workload, stress etc.

An important aspect of this work will be to explore the extent to which trust in automation influences choices that are made in the course of working with automation. Within the interpersonal trust literature, proposed benefit of trust is that it enables "risky" choices, that it makes one more able to take a "leap of faith" into the unknown. In a military context, trust in automation may influence critical decisions that put the automation more in control than if trust were low. Again, this issue has yet to be explored in the existing trust in automation research, and has not been explored within the military context.

The potential dangers of trust in automation are an especially important area to investigate in a research program, as this issue has serious implications in military environments. The goal in this stage would be to establish what the optimal level of trust in automation should be, given the context in which a trust judgement is made.

The remainder of this chapter outlines the features of a potential research program, offers research questions that might be addressed in the study of trust in automation, and proposes a research approach based on a prototype study.

11.3 Features of a Research Program

Task-based. Different kinds of tasks will have different implications for trust in automation. Tasks that require a high level of complex interaction with automation, for example, are more likely to require higher levels of trust in the automation than more simplistic tasks. It is important, then, to clearly define the task within any future research program, and to control for task difficulty and complexity in understanding trust in automation.

Mission-based. Trust in automation should be explored in the context of missions that require the use of automation. An important aspect of this research will be anticipating the different challenges to trust in automation presented during the course of missions. The manipulation of risk in various kinds of missions is one obvious way in which the contextual factors evident in actual missions can be represented. In addition, in light of the unique nature of the military domain, it will be important to consider trust in automation in mission scenarios with and without adversaries. Scenarios with adversaries present unique challenges to trust in automation, as tampering and intentional sabotage has the potential to disrupt not only the reliability of the automation, but the trust-relevant attitudes toward it.

Organization-based. Understanding trust in automation cannot be done in isolation of the organizational context in which it occurs. As noted earlier, the rules and procedures in place to govern interaction with automation are likely to have a serious impact both on how automation is used, and on the extent to which it is trusted. As such, it is important to understand organizational factors as well, and trust in automation needs to be considered separately from organizational mandates or demands to use automation.

Methodological Issues

Certain control conditions need to be incorporated into any research program. For the study of trust in automation, the following issues are seen as the most important:

Level of Automation. As noted throughout this report, with so many different forms of automation, it is unclear whether the factors that influence trust are necessarily similar. For example, issues associated with trusting a fairly simple but mechanistic form of automation are the same as those associated with trusting the recommendation of a complex decision aid. It may be that the factors influencing trust in these two different systems are somewhat different in magnitude. In any case, this is an empirical question that needs to be resolved in future research.

Experimental Tasks. Experimental tasks and mission scenarios need to be carefully planned. Trust in automation develops over time and rests on the opportunity for interactions. It is critical that at least some parts of the research program employ a longitudinal approach and explore the gradual accumulation of trust over repeated interactions with automation. The use of time series design with an even longer time frame than was used in previous trust in automation research may assist in understanding trust in automation at a truly dynamic level.

Secondly, it will be important to find a balance between maximizing the experimental realism of the military setting, and exerting the necessary experimental controls. Although it is important that the proper controls are in place, it is critical that the true essence of trust is represented in the experimental tasks. At the early stages of exploring trust automation, however, there may be some merit in conducting work with higher control but less realism, and there may be value in beginning research in more artificial and controlled settings. Once the phenomenon of trust in automation has been captured, this would allow expansion to more realistic settings.

The experimental study of trust in automation may require creating situations in which the contextual factors of risk, vulnerability, and uncertainty are prominent. As we noted in a previous review (Adams & Webb, 2003), there are ethical issues inherent with creating high levels of risk and uncertainty in experimental settings. In the existing trust in automation research, risk has been manipulated at a very minor level, by increasing the speed at which a process control task needed to be performed (Lewandowsky et al., 2000). In order to understand trust in automation within a military context, ultimately, risk manipulations may need to be much more realistic in order to parallel the pace and stress of operations. This will need to done with care in order to ensure that participants are not harmed. It is important, for example, in creating automated system faults that attitudes and expectations about automation in general are not likely to generalize and to come to influence attitudes and behaviour outside of the experimental context.

Specific Research Questions

We believe that the following issues represent the most pressing challenges to understanding trust in automation, and to applying this knowledge to the broader study of trust in automation. Research exploring trust in automation should focus on the following three questions.

Table 18: Questions related to the nature of trust in automation

Research Goal: Determine the nature of trust in automation	
Research issue	Example Research Question
What is the structure of trust in automation?	Is trust in automation determined by both cognitive and affective factors? Can person-based and category-based trust in automation be distinguished? Is trust in automation best represented as a single bipolar trust/distrust construct or as trust and distrust separately?
What function does trust in automation serve?	Does trust enable prediction of automation? Does trust in automation promote positive attributions about the automation? Does trust in automation reduce uncertainty, risk, and vulnerability?

Table 19: Questions related to the factors that develop and maintain trust in automation.

Research Goal: Determine the factors that develop and maintain trust in automation	
Research issue	Example Research Question
What factors related to the properties of automation affect the development of trust in automation?	How do the following properties of automation affect trust in automation: • System reliability • System faults • System output and feedback • Explication of intention • Interactivity • Ability to form mental models • Susceptibility to tampering
What factors related to the properties of the trustor affect the development of trust in automation?	How do the following properties of the trustor affect trust in automation: • Self-confidence • Propensity to trust • Trust history • Age • Culture
What contextual factors affect the development of trust in automation?	How do the following contextual factors affect trust in automation: • Operational context • Organizational factors • Task factors • Training

Table 20: Questions related to how trust in automation impacts on performance.

Research goal: Determine how trust in automation impacts on performance.	
Research Issue	Example Research Questions
How is defensive monitoring affected by trust in automation?	Do high trust operators monitor automation less?
How is the use of automation affected by trust in automation?	Do high trust operators use automation more?
How is overall job performance affected by trust in automation?	Do high trust operators perform their jobs more effectively and efficiently?
Can performance be predicted by knowledge of trust in automation?	Can measured increases in trust in automation be shown to impact incrementally on performance?

11.4 Proposed Research Approach

The next section defines the steps of a research program that could be used to study trust in automation.

Establish mission types of interest. Consideration must be given to the types of operations to be studied in the research project. This decision will be determined, in part, by the research questions to be explored. Simulated conventional combat operations, for example, provide ideal settings within which to study issues of how contextual factors such as risk and uncertainty are likely to impact on the development of trust in automation, and to explore how the relationship between trust and performance.

Establish command levels of interest. As this project's focus is on trust in automation, and as automation is used widely across varying levels (e.g. command and control), multiple command levels could potentially come into play. In designing the current research, then, it will be important consider the possibility that future research may require more complex command levels and, to the greatest degree possible, to structure this research to ensure maximal flexibility.

Establish a mission context (scenario). It is important to decide on a scenario or a mission context that will define the goals of the mission, the resources available and the environment in which the mission will occur. Decisions about mission contexts for any given study will obviously be determined by the questions that the research addresses. In understanding that factors that influence trust in automation, for example, it will be important to establish a mission context that will enable the creation of varying levels of risk and uncertainty, in order to understand the influence of various properties of the automation. Once trust in automation is established, and the question turns to the impact of trust in automation, a different kind of context will be required. At this point, it will be important to be able to establish a context in which organizational factors that typically influence trust in automation are either controlled or manipulated in order to explore the relationship between trust as a psychological state and the use of automation. As such, creating mission contexts in which these questions can be explored will be important.

Establish the type and level of automation. Automation varies widely in terms of its complexity. We have argued that the dynamics of trust with very complex forms of automation may not be the case as those with more simple forms of automation. Similarly, it is also possible that issues of trust are somewhat different when the output from automation is more physical vs. when the automation gives advice to a human. It will be important, then, to clearly delineate the kind of automation that will be researched at the earliest possible stage so that the research approach can be best tailored to the type of automation.

Establish study sampling frame. The population from which participants are chosen during the course of the research program will likely shift as the research program proceeds. Early studies may allow for less constrained samples than later studies. At the earliest stages of research, for example, it may be possible to work with novice operators in order to establish basic principles, but highly experienced operators may be critical to work with at the later stages of research.

Establish an appropriate simulation environment. Trust in automation can be meaningfully studied in many different ways. These include somewhat more artificial laboratory studies.

The problem with the existing research, however, is that the results sometimes seem to be at least somewhat specific to the kind of microworld in which the research was conducted. This may have hindered the efforts of researchers in establishing a widely applicable range of basic principles about what trust in automation is like. As such, much of the empirical literature speaks to findings that because of the context in which they were studied are only applicable and valid with many qualifications. In this sense, going beyond the one kind of paradigm that has typically been used is critical.

In this sense, the ability to use simulation in order to parallel the kinds of decisions required in command and control contexts, for example, will be critical. These simulations should parallel the battlefield environment, and require multiple sources of information from both known and relatively less known data sources. A simulated decision aid will also be required in order to fuse this information together and make recommendations as to the optimal course of action. Such a simulation must also be able to systematically vary different kinds of faults (e.g. intentional vs. unintentional, and timing (e.g. constant, intermittent or random) of the fault. The system must also be designed in such a way that the decision making processes in response to the output given by the decision aid allow for multiple possible responses (rather than a simple binary choice such as use or non-use of the information). Most importantly, trust as a psychological state and the related courses of action that arise need to be measurable.

Paradigms in which the advice actually provided by an automated decision aid is purported to be from either a human or a decision aid (e.g. Lewandowsky et al., 2000) are potentially important tools that allow for identifying what is unique about our interactions with automation vs. trust in human relationships. Moreover, this kind of paradigm also holds considerable promise in the sense that it allows the exploration of attributions about the automation.

A more realistic field study could be conducted in the context of actual work with automation. The experience of operators working with a new piece of automation, for example, could be tracked, and data relevant to their trust in the automation and related factors could be probed at either at regular intervals or after specific system events (e.g. automation faults) occur. In this scenario, a data collection device (e.g. PDA) synchronized with automation performance or otherwise incorporated into interaction with automation could enable data collection. Within this kind of experimental setting, of course, it is impossible to control contextual factors such as risk.

Determining what form of simulation to use has important implications for the validity of the results to be gained in each of these settings. The various forms of simulation noted vary widely in terms of their fidelity to trust in real life infantry situations, and have both advantages and disadvantages. Low fidelity simulations, such as those occurring in strictly controlled experiments, can be used with relatively little effort and offer the opportunity for clear and concise data measurement. Whether the results are valid, however, depends on how well the simulation truly captures trust. Simulations with a higher level of fidelity require more elaborate preparation and resources, and getting data in more realistic settings tends to be harder. Higher fidelity simulations, however, present fewer challenges to validity. In studying trust, the issue of how well simulations of trust-relevant situations mirror trust in automation globally is an important issue to consider in terms of both validity and ease of implementation.

As such, it seems important to explore some of the key issues already evident in the literature. One of the most interesting issues, for example, is the finding that when people have advance knowledge of system faults, these faults do not necessarily diminish trust. The problem with this interesting finding, however, is that it has never been directly compared in the context of an experiment that systematically presented a system with faults in which people did not have advance knowledge versus a control condition in which people did have advance knowledge of faults. As such, it is unclear whether this finding would hold under closer scrutiny. A prototypical study that might help to refine this finding is described below:

11.5 Prototypical Study

Purpose – To explore the impact of advance knowledge of faults on trust in automation.

Outline – A computer administered scenario could be conducted in a military training simulation. The scenario would have three distinct stages, representing pre-assessment, simulation and post-assessment. The task is to make decisions regarding whether targets are friendly, hostile or unknown, based on the recommendations presented by the decision aid.

Method – In this scenario, a participant's task is to identify the status of targets on a mapped radar display. The goal of the mission is to make the best choices in accordance with the information presented.

This experiment can be conceptualized as having three stages. Before the task-related stage of the simulation begins, a number of dimensions relevant to trust in automation would be measured. These dimensions include propensity to trust automation in general, and trust in the specific automation (even before interacting with it). In addition, knowledge surveys exploring participant's expectations about the automation would also be completed by participants, as well as their predictions of success at the task. Observers would also rate the success of the mission in terms of task effectiveness and task efficiency, as well as objective measures (e.g., time to task completion).

Information relevant to the potential problems of the decision aid will then be presented to some participants. In essence, these participants will receive detailed information about the ways in which the decision aid is likely to fail. Participants in the control condition will receive information relevant to the decision aid, but which is non-diagnostic with respect to the future performance of the decision aid.

After these assessments, the simulation then becomes active, and the participant works with the decision aid to complete the assigned task. Ongoing measures of performance (e.g. efficiency identifying targets, time to make decision) and defensive monitoring of the automation are taken. At regular intervals, the system will also prompt for ratings of trust in the automation, using the measure of trust in automation described earlier, and similar freeze probe techniques would be used to assess the degree to which the automation is performing in accordance with one's prior expectations. Updated predictions about the probable success of the task will be taken. At predefined intervals, the automated decision aid will be programmed to emit errors. The simulation will then resume, and similar ongoing measures of process and performance are taken.

Once the task is complete or criteria time reached, measures of trust in automation would again be taken, as well as post scenario study of how the human operators will be able to predict the automation's behaviour during the task and the success of the "mission". Successive iterations of this scenario would indicate the changes in trust over the course of time, and as the result of advance knowledge of automation faults. As a whole, then, analyses on the data would explore changes in trust and related indicators over the course of the simulation.

Additional Variations – Many different variations of this simulation are possible. In addition to trust in automation with and without advance knowledge of faults, it would also be possible to explore varying kinds of faults at a deep level and at a more surface level, in addition to varying sizes and magnitude of faults. An additional variation could be a scenario that imposes an unforeseen circumstance during the automation's functioning. Another possible variation would be to alter the transparency of the automation by making the decision rules that the automation uses explicit. Another possible variation would be to vary the perceived level of risk.

Possible Hypotheses – Possible hypotheses might include:

Operators with advance knowledge of faults will experience less decrease in their trust once these errors occur than will operators with no such knowledge

Operators with advance knowledge of faults will show less defensive monitoring

Operators with advance knowledge of faults will be more proficient at accurately estimating their performance than will operators with no advance knowledge of faults

Anticipated Lessons Learned – This study provides a good test of the relationship between trust and advanced knowledge of faults.

Limitations – This study requires considerable programming and resources. The freeze probe disrupts the flow of the mission.

PRIMARY REFERENCES

BISANTZ, A., LLINAS, J., SEONG, Y., FINGER, R. and JIAN, J. (2000). Empirical investigations of trust-related systems vulnerabilities in aided, adversarial decision making – Phase 3. Wright-Patterson AFB, OH, Air Force Research Laboratory: 95.

COHEN, M., PARASUMARAN, R., and FREEMAN, J. (1998). Trust In Decision Aids: A Model And Its Training Implications. Proceedings, 1998 Command and Control Research and Technology Symposium. Department of Defense C4ISSR Cooperative Research Program.

DIJKSTRA, J. J. (1999). User agreement with incorrect expert system advice. Behaviour & Information Technology, 18(6): 399-411.

ENDSLEY, M.R. and KIRIS, E.O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors*, 37, 381-394.

GOILLAU, P., KELLY, C., BOARDMAN, M. and JEANNOT, E. (2001). Development of a measure of trust in ATM systems. European Organisation for the Safety of Air Navigation: 60pp.

JIAN, J., BISANTZ, A. and DRURY, C. (2000). Foundations for an empirically determined scale of trust in automated systems. International Journal of Cognitive Ergonomics, 4(1), 53-71.

KELLY, C., BOARDMAN, M., GOILLAU, P. and JEANNOT, E. (2001). Principles and Guidelines for the Development of Trust in Future ATM Systems: A Literature Review, European Organisation for the Safety of Air Navigation: 48pp.

LEE, J. and MORAY, N. (1992). Trust, control strategies and allocation of function in human-machine systems. Ergonomics, 35(10), 1243-1270.

LEE, J. and MORAY, N. (1994). Trust, self-confidence, and operators' adaptation to automation. International Journal of Human-Computer Studies, 40, 153-184.

LERCH, F, PRIETULA, M. and KULIK, C. (1997). The turing effect: the nature of trust in expert system advice. Expertise in Context: Human and Machine. P. J. Feltovich and K. M. Ford (Eds.), MIT Press, 417-448.

LEWANDOWSKY, S., MUNDY, M. and TAN, G. (2000). The dynamics of trust: Comparing humans to automation. Journal of Experimental Psychology: Applied, 6(2), 104-123.

LLINAS, J., BISANTZ, A., DRURY, C., SEONG, Y. and JIAN, J. (1998). Studies and analyses of aided adversarial decision making. Phase 2: Research of human trust in automation. Wright-Patterson AFB, OH, Air Force Research Laboratory, Human Effectiveness Directorate: 117.

LLINAS, J., DRURY, C., BIALAS, W. and CHEN, A. (1998). Studies and analyses of vulnerabilities in aided adversarial decision making - Phase I. Wright-Patterson AFB, OH, Air Force Research Laboratory, Human Effectiveness Directorate: 123.

MADSEN, M. and GREGOR, S. (2000). Measuring Human-Computer Trust. Gladstone, Australia, Central Queensland University: 12pp.

MORAY, N., HISKES, D., LEE, J. and MUIR, B. (1995). Trust and human intervention in automated systems. In Expertise and technology : Cognition and human-computer cooperation (pp. 183-194). Hillsdale, NJ : Lawrence Erlbaum Associates.

MORAY, N. and INAGAKI, T. (1999). Laboratory studies of trust between humans and machines in automated systems. Trans Inst MC 21(4/5): 203-211.

MORAY, N., INAGAKI, T. and ITOH, M. (2000). Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. Journal of Experimental Psychology: Applied, 6(1), 44-58.

MUIR, B. M. (1994). Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems, Ergonomics 37(11), 1905-1922.

MUIR, B. M. and MORAY, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. Ergonomics, 39(3), 429-460.

PARASURAMAN, R. and RILEY, V. (1997). Humans and automation: Use, misuse, disuse, abuse. Human Factors, 39(2), 230-253.

RILEY, V. A. (1994). Human use of automation. Ph.D. Thesis submitted to the University of Minneapolis.

SEONG, Y. and BISANTZ, A. M. (2000). Modeling human trust in complex, automated systems using a lens model approach: 95-100.

SHERIDAN, T. B. (1988). Trustworthiness of command and control systems. IFAC Man-Machine Systems: 427-431.

SHERIDAN. T. B. (2002). Humans and automation: System design and research issues. HFES Issues in Human Factors and Ergonomics Series, Volume 3. Santa Monica, CA: Wiley and Sons Publications, Inc.

SINGH, I., MOLLOY, R. and PARASUMARAN, R. (1993). Automation-induced complacency: Development of the complacency-potential rating scale. The International Journal of Aviation Psychology, 3(2), 111-122.

TYLER, R. R. (1999). Human automation interaction - a military user's perspective. In Automation Technology and Human Performance. M. W. Scerbo and M. Mouloua (Eds.). Mahwah, NJ, Lawrence Erlbaum Associates, 38-41.

SECONDARY REFERENCES

ADAMS, B. D. & WEBB, R. D. G. (2003). Model of trust development in small teams. *Report to Defence and Research Development Canada- Toronto*. Humansystems Inc. Guelph, Ontario, Canada. Technical report CR-2003-016.

ADAMS, B. D., BRYANT, D. J., & WEBB, R. D. G. (2001). Trust in teams: Literature review. *Report to Defence and Civil Institute of Environmental Medicine*. Humansystems Inc. Guelph, Ontario, Canada. Technical report CR-2001-042.

ALCOCK, J.E., CARMENT, D.W., & SADAVA, S.W. (2001). A Textbook of Social Psychology (5th ed). Scarborough, Ontario: Prentice-Hall Canada.

ALDERN, T. (1995). Intelligent system operational support requirements. In the Human/Electronic Crew: Can we trust the team? Proceedings of the 3rd International Workshop of Human-Computer Teamwork.

BARBER, B. (1983). The logic and limits of trust. Rutgers University Press, New Brunswick.

BEM, D. (1967). Self Perception: An Alternative Interpretation of Cognitive Dissonance Phenomena, *Psychological Review*, 74(3), 183-200.

BOON, S., & HOLMES, J. (1991). The dynamics of interpersonal trust: Resolving uncertainty in the face of risk. In Hindle, R., & Groebel, J. (Eds.). *Cooperation and Prosocial Behavior*, (pp.167-182). New York: Cambridge University Press.

BROWER, H., SCHOORMAN, F. and TAN, H. (2000). A Model of Relational Leadership: The Integration of Trust and Leader-Member Exchange. Leadership Quarterly 11(2): 227-250.

BRUNSWIK, E. (1952). The conceptual framework of psychology. (International Encyclopedia of Unified Science, Volume 1, Number 10.) Chicago: The University of Chicago Press.

CHALMERS, B.A., EASTER, J.R., and POTTER, S.S. (2000). Decision-Centred Visualisations for Tactical Decision Support on a Modern Frigate. Proceedings of the Command and Control Research Technology Symposium.

COHEN, M.S. (2000). A Situation Specific Model of Trust in Decision Aids. Proceedings of the International Conference on Human Performance, Situation Awareness & Automation. Savannah, Ga.

COSTA, A. C., ROE, R. A. and THAILLEAU, T. (2001). Trust Within Teams: The Relation With Performance Effectiveness, European Journal of Work and Organizational Psychology, 10(3): 225-244.

DIRKS, K. T. and FERRIN, D. (2002). Trust in Leadership: Meta-Analytic Findings and Implications for Research and Practice. Journal of Applied Psychology, 87(4).

DOUGHERTY, S. (2003). Automation. Retrieved Feb. 5, 2003 from http://faculty.erau.edu/dohertys/410/410_16_automation.ppt

FITTS, P. H. (1951). Human engineering for an effective air navigation and traffic control system. Washington, DC: National Research Council.

HOSMER, L. (1995). Trust: The connecting link between organizational theory and philosophical ethics. *Academy of Management Review*, 20(2), 379-403.

HUTCHINS, S. G. (1996). Principles for intelligent decision aiding. Arlington, VA, Office of Naval Research: 34.

JONES, G. and GEORGE, J. (1998). The experience and evolution of trust: Implications for cooperation and teamwork. *Academy of Management Review*, 23(3). 531-546.

KRAMER, R.M. (1999). Trust and distrust in organizations: Emerging perspectives, enduring questions. *Annual Review of Psychology, 50*, 569-598

LEVIS, A.H., MORAY, N., & HU, B.S. (1994). Task decomposition, and allocation problems and discrete event systems. *Automatica, 30*, 203-216.

LEWICKI, R., & BUNKER, B. (1996). Developing and maintaining trust in work relationships. In Kramer, R. & Tyler, T. (Eds.). *Trust in organizations: Frontiers of theory and research.* (pp. 114-139). Thousand Oaks, CA, US: Sage Publications, Inc.

LEWICKI, R., MCALLISTER, D., & BIES, R. (1998). Trust and distrust: New relationships and realities. *Academy of Management Review*, 23(3), 438-445.

LEWIS, J. D. and WEINGERT, A. J. (1985). Trust as a social reality. *Social Forces*, 63, 967-985.

LUHMANN, N. (1988). Familiarity, confidence, trust: Problems and alternatives. In D. Gambetta (Ed.). *Trust: Making and breaking cooperative relations*. (pp. 94-108). New York: Basil Blackwell.

MASTAGLIO, T. (1999). Automation and human performance: Consideration in future military forces. In M.W. Scerbo & M. Mouloua (Eds.), *Automation Technology and Human Performance: Current Research and Trends.* Lawrence Erlbaum Associates: Mahwah, N.J., pp. 38-41.

MATTHEWS, M.L., WEBB, R.D.G., BRYANT, D.J., 1999. Cognitive Task Analysis of the HALIFAX-Class Operations Room Officer. *Report to Department of National Defence.*

MAYER, R., DAVIS, J., and SCHOORMAN, F. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20(3), 709-734.

MCALLISTER, D. (1995). Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, 38(1), 24-59.

MCFADDEN, S.M. GIESBRECHT, B.L., GULA, C.A. (1998). Use of an automatic tracker as a function of its reliability. *Ergonomics,* 41(4), 512-536.

MOLDAFSKY, N.I., & KWON, I.-W. (1994). Attributes affecting computer-aided decision making - a literature survey. *Computers in Human Behavior,* 10(3), 299-323.

MUIR, B. M. (1989). Operator's trust in and use of automatic controllers in a supervisory process control task. Unpublished doctoral dissertation, University of Toronto.

NORTHRUP, B. (2002). Sea King Status – An Aviator's Perspective. Retrieved Feb. 20, 2003 from http://www.naval.ca/article/Northrup/Sea_King_Status.pdf.

OMODEI, M. and MCLENNAN, J. (2000). Conceptualizing and measuring global interpersonal mistrust-trust. *Journal of Social Psychology*, 140(3), 279-294.

ONKEN, R. (1999). The cognitive cockpit assistant systems CASSY/CAMA. SAE Technical Paper Series: 1-7.

REMPEL, J.K., HOLMES, J.G., and ZANNA, M.P. (1985). Trust in close relationships. *Journal of Personality and Social Psychology, 49*, 95-112.

RILEY, V. (1996). Operator reliance on automation: Theory and data. In R. Parasumaran and M. Mouloua (Eds.), Automation and human performance: Theory and applications (pp. 19-35). Hillsdale, NJ: Erlbaum.

ROTTER, J. (1967). A New Scale For The Measurement Of Interpersonal Trust. *Journal of Personality*, 35(4), 651-665.

SARTER, N. B. and WOODS, D. (1997). Team play with a powerful and independent agent: operational experiences and automation surprises on the Airbus A-320." Human Factors 39(4): 553-569.

SHAW, R.B. (1997). Trust in the Balance: Building successful organizations on results, integrity, and concern. Jossey-Bass: San Francisco.

TAYLOR, R.M., HOWELLS, H. and WATSON, D. (2000). The cognitive cockpit: operational requirement and technical challenge. In P.T. McCabe, M.A. Hanson, & S.A. Robertson (Eds.), *Contemporary Ergonomics 2000*, 55-59. Taylor and Francis: London.

ZUBOFF, S. (1988). In the age of the smart machine: The future of work and power. Basic: New York.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A – SCALE OF TRUST IN AUTOMATED SYSTEMS

Checklist for Trust between People and Automation

Below is a list of statement for evaluating trust between people and automation. There are several scales for you to rate intensity of your feeling of trust, or your impression of the system while operating a machine. Please mark an "x" on each line at the point which best describes your feeling or your impression. .[10]

(Note: not at all=1; extremely=7)

1 The system is deceptive

1 2 3 4 5 6 7

2 The system behaves in an underhanded manner

1 2 3 4 5 6 7

3 I am suspicious of the system's intent, action, or outputs

1 2 3 4 5 6 7

4 I am wary of the system

1 2 3 4 5 6 7

5 The system's actions will have a harmful or injurious outcome

1 2 3 4 5 6 7

---

[10] This checklist matches the original exactly, although it is unclear whether the authors intended for the numbers and the rating scale to be misaligned.

6 I am confident in the system

1 2 3 4 5 6 7

7 The system provides security

1 2 3 4 5 6 7

8 The system has integrity

1 2 3 4 5 6 7

9 The system is dependable

1 2 3 4 5 6 7

10 The system is reliable

1 2 3 4 5 6 7

11 I can trust the system

1 2 3 4 5 6 7

12 I am familiar with the system

1 2 3 4 5 6 7

(In Jian, Bisantz and Drury, 2000)

APPENDIX B – COMPLACENCY-POTENTIAL RATING SCALE

Instructions

Read each statement carefully and check one response out of five alternatives in the appropriate box which you feel accurately described your views or experiences. The responses vary on a scale of agreement / disagreement, from "strongly agree" to "strongly disagree:. For example:

Statement:

Doing research in a library has been made easier by the introduction of computerized card cataloguing systems.

☐ Strongly agree ☐ Agree ☐ Undecided ☐ Disagree ☐ Strongly disagree

Give your answer for each statement and be sure to place your response in the correct place. Remember, this is an opinion survey and not a test of intelligence or ability. There are no right or wrong answers, only answer that fit your views accurately. Do not skip any question. Time is limited. Do you have any questions?

1. Manually sorting through card catalogues is more reliable than computer-aided searches for finding items in a library.
2. If I need to have a tumour in my body removed, I would choose to undergo computer-aided surgery using laser technology because computerized surgery is more reliable and safer than manual surgery.
3. People save time by using automatic teller machines (ATMs) rather than a bank teller for banking transactions.
4. I do not trust automated devices such as ATMs and computerized airline reservation systems.
5. People who work frequently with automated devices have lower job satisfaction because they feel less involved in their job than those who work manually.
6. I feel safer depositing my money at ATM than with a human teller.
7. I have to tape an important TV program for a class assignment. To ensure that the correct program is recorded, I would use the automatic programming facility on my VCR rather than manually taping.
8. People whose job requires them to work with automated systems are lonelier than people who do not have work with such devices.
9. Automated systems used in modern aircraft, such as the automatic landing system, have made air journeys safer.

10. ATMs provide a safeguard against the inappropriate use of an individual's bank account by dishonest people.
11. Automated devices used in aviation and banking have made work easier for both employees and customers.
12. I often use automated devices.
13. People who work with automated devices have greater job satisfaction because they feel more involved than those who work manually.
14. Automated devices in medicine save time and money in the diagnosis and treatment of disease.
15. Even though the automatic cruise control in my car is set at a speed below the speed limit, I worry when I pass a police radar speed-trap in case the automatic control is not working properly.
16. Bank transactions have become safer with the introduction of computer technology for the transfer of funds.
17. I would rather purchase an item using a computer than have to deal with a sale representative on the phone because my order is more likely to be correct using the computer.
18. Work has become more difficult with the increase of automation in aviation and banking.
19. I do not like to use ATMs because I feel that they are sometimes unreliable.
20. I think that automated devices used in medicine, such as CAT-scans and ultrasound, provide very reliable medical diagnosis.

APPENDIX C - SHAPE Automation Trust Index (SATI v0.3)

SATI Part 1 (*please complete before the start of the day's simulation runs*)

Please tell us who you are, and your forthcoming role in the simulation. Thank you.

About you:

Name:	
Nationality:	
Sex (M /F):	

About the simulation:

Date and time:	
Name of simulation project:	
Computer-assistance or automation tools available:	1. 2. 3. 4. 5.
Your simulated sector:	
Your role (Planner / Executive controller)	

SATI Part 1 (*continued*)

PLEASE COMPLETE AT THE START OF EACH DAY

1. What do you think of the simulation so far? (Please mark the scale with an 'X').

Bad	**OK**	**Good**

2. Are you prepared to trust the simulated system? Please give your reasons.

No	**Yes**

3. How much confidence do you have in the simulated system? (Please mark the scale with an 'X').

None	**OK**	**Full**
0%	**50%**	**100%**

4. Please give your reasons

SATI Part 2 (*please complete after the end of the simulation runs*)

Please write your name and your last role in the simulation. Thank you.

About you:

Name	

About the simulation:

Date and time:	
Name of simulation project:	
Computer-assistance or automation tools available:	1. 2. 3. 4. 5.
Your last simulated sector:	
Your last role (Planner / Executive controller)	

SATI Part 2 (*continued*)

PLEASE COMPLETE AT THE END OF THE DAY'S RUNS

Based on today's runs

1. What did you think of the simulation? (Please mark the scale with an 'X').

Bad	**OK**	**Good**

2. Were you prepared to trust the simulated system?

No	**Yes**

3. How much confidence did you have in the simulated system? (Please mark the scale with an 'X').

None	**OK**	**Full**
0%	**50%**	**100%**

4. Please give your reasons. If your trust or level of confidence in the system has changed since the start of the day, please explain why.

SATI Part 2 (*continued*)

PLEASE COMPLETE A SEPARATE SHEET FOR EACH AVAILABLE AUTOMATION TOOL.

5. Please judge each automation tool against the following factors (mark each scale with an 'X').

Name of automation tool: ______________________________

1. Is the automation tool useful ?

☹	*Not useful*	-5 ┼─┼─┼─┼─┼─┼	0	┼─┼─┼─┼─┼─┼ +5	*Useful*	☺

2. How reliable is it ?

☹	*Not reliable*	-5 ┼─┼─┼─┼─┼─┼	0	┼─┼─┼─┼─┼─┼ +5	*Reliable*	☺

3. How accurately does it work ?

☹	*Not accurate*	-5 ┼─┼─┼─┼─┼─┼	0	┼─┼─┼─┼─┼─┼ +5	*Accurate*	☺

4. Can you understand how it works ?

☹	*Not understand*	-5 ┼─┼─┼─┼─┼─┼	0	┼─┼─┼─┼─┼─┼ +5	*Understand*	☺

5. Do you like using it ?

☹	*Dislike*	-5 ┼─┼─┼─┼─┼─┼	0	┼─┼─┼─┼─┼─┼ +5	*Like*	☺

6. How easy is it to use ?

☹	*Difficult*	-5 ┼─┼─┼─┼─┼─┼	0	┼─┼─┼─┼─┼─┼ +5	*Easy*	☺

6. Please rank these factors in order of relative importance. Number them from 1 (*least important)* to 6 (*most important*). Please use each number once only.

NAME OF AUTOMATION TOOL:________________	
Usefulness	*ranking:*
Reliability	*ranking:*
Accuracy	*ranking:*
Understanding	*ranking:*
Liking	*ranking:*
Ease of use	*ranking:*

SATI Part 2 (*continued*)

LOOKING BACK OVER THE DAY'S SIMULATION RUNS:

7. Please rate your amount of confidence in each of these five dimensions.

Please mark each scale with an 'X'.

1. Confidence in automation tools

0 50 100 %

2. Confidence in simulation

0 50 100 %

3. Self-confidence

0 50 100 %

4. Confidence in controller colleagues

0 50 100 %

5. Confidence in pilots

0 50 100 %

8. Would you work live traffic with the tools? In your opinion, what changes would the automation need so that your trust and confidence would be increased?

   If there are any other factors which influence your trust in an ATC system, or if you have any general comments, please write them here.

Thank you for completing this questionnaire.

DOCUMENT CONTROL DATA SHEET

1a. PERFORMING AGENCY

Humansystems, Incorporated, 111 Farquhar St., 2nd floor, Guelph, ON N1H 3N4

2. SECURITY CLASSIFICATION

UNCLASSIFIED

-

1b. PUBLISHING AGENCY

DRDC Toronto

3. TITLE

(U) TRUST IN AUTOMATED SYSTEMS: LITERATURE REVIEW

4. AUTHORS

Barbara D. Adams; Lora E. Bruyn; Sébastien Houde

5. DATE OF PUBLICATION

June 23 , 2003

6. NO. OF PAGES

136

7. DESCRIPTIVE NOTES

8. SPONSORING/MONITORING/CONTRACTING/TASKING AGENCY

Sponsoring Agency:

Monitoring Agency:

Contracting Agency : DRDC Toronto

Tasking Agency:

9. ORIGINATORS DOCUMENT NO.

Contract Report CR 2003-096

10. CONTRACT GRANT AND/OR PROJECT NO.

W7711-017747/001/TOR Call-Up 7747-10

11. OTHER DOCUMENT NOS.

12. DOCUMENT RELEASABILITY

Unlimited distribution

13. DOCUMENT ANNOUNCEMENT

Unlimited announcement

14. ABSTRACT

(U) This report reviews research literature pertaining to trust in automated systems. Based on the review, we argue that trust in automation has many similarities with trust in the interpersonal domain, but also several unique dynamics and influences. Existing research has focused primarily on trust in automation that has an executive or control function, and to a lesser extent, has considered trust in automation that is designed to present information to operators (e.g. decision aids). We maintain that although there are many similarities between trust in automation and interpersonal trust, the dynamics of trust in automation also have some distinct qualities. Several models related to trust in automation have already been developed; in this report, a comprehensive -- although still preliminary -- model of trust in military automation is proposed. Several sets of factors are likely to impact on the development of trust in automation, including properties of the automation, properties of the operator, and properties of the context in which interaction with automation occurs. The consequences of trust in automation have yet to be fully explored. Based on this review, measures and methods to study trust in automation are considered, and a program of research to study trust in automated systems is described.

(U) Ce rapport examine les publications dans le domaine de la recherche sur la confiance qui peut être accordée aux systèmes automatisés. Cette revue nous amène à conclure que la confiance dans l'automatisation présente de nombreuses similitudes avec la confiance dans le domaine interpersonnel, mais qu'elle présente aussi d'uniques dynamiques et subit des influences particulières. Jusqu'à présent, la recherche s'est axée principalement sur la confiance dans l'automatisation des fonctions de supervision ou de commande et, dans une moindre mesure, sur l'automatisation conçue en vue de présenter de l'information à des opérateurs (p. ex. les aides à la décision). Nous sommes d'avis que, bien qu'il y ait de nombreuses similitudes entre la confiance dans l'automatisation et la confiance dans le domaine interpersonnel, la dynamique de la première a des qualités distinctives en sus. Plusieurs modèles ont déjà été développés dans le domaine de la confiance dans l'automatisation et le présent rapport en propose un qui est complet et qui s'applique au contexte militaire. Plusieurs ensembles de facteurs sont susceptibles d'influer sur le développement de la confiance dans l'automatisation, notamment les propriétés de l'automatisation, les propriétés de l'opérateur et les propriétés du contexte dans lequel a lieu l'interaction avec l'automatisation. Les conséquences de la confiance dans l'automatisation demeurent à explorer plus avant. Cette revue des publications a servi de base à l'examen de mesures et méthodes d'étude de la confiance dans l'automatisation et un programme de recherche en vue d'étudier la confiance dans les systèmes automatisés est décrit.

15. KEYWORDS, DESCRIPTORS or IDENTIFIERS

(U) literature review; trust; automated systems; military; model